Standalone LLM Prompts

1. Bubble Only Configuration

1-1. Standalone LLM Prompt (Bubble-Only Fed with Bubble Data)

Act as an impartial, closed-world evaluator: using only the de-identified natural-language summaries of six U.S. macro-financial indicators provided below—without external or historical knowledge—assign a probability that the target 24-month window is a bubble by strictly following the rules specified in the continuing sections.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat the reference as an anonymized, widely recognized bubble prototype; treat the target as an anonymized window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- All data describe the United States equity market and U.S. macro-financial indicators. The target is an anonymized 24-month U.S. window; the reference is an anonymized, widely recognized U.S. bubble prototype.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the NL summaries of the reference and target files.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A reference bubble prototypes represented as natural-language summaries for each of six indicators.

- A target 24-month window is represented the same way. This target window is based on an unknown period.

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

4) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

5) Your task (free qualitative reasoning):

- Inputs to use: the de-identified NL summaries for the Target 24-month window and the Reference bubble prototype (pattern-level, anonymized).

- Goal: Assign a single probability that the Target 24-month window is in a bubble state, using only the information provided here.

- Method: Read the provided summaries carefully and compare them for the stated goal. Do not rely on any external knowledge, numeric thresholds, or unstated calibration rules.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so explicitly and avoid confident extremes.

6) Output (exactly this format):

- Probability: <a single number in [0,1] with four decimals representing P(bubble).

- Rationale: Justify your decision based only on the provided natural summaries of the macro-financial indicators of the references.

7) Data (de-identified NL summaries):

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The bubble data has tag “Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here:

{target}

{Bubble_Prototypes}

1-2. LLM 프롬프트 (Bubble-Only Fed with Non-Bubble Data)

Act as an impartial, closed-world evaluator: using only the de-identified natural-language summaries of six U.S. macro-financial indicators provided below—without external or historical knowledge—assign a probability that the target 24-month window is a bubble by strictly following the rules specified in the continuing sections.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat the reference as an anonymized, widely recognized bubble prototype; treat the target as an anonymized window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- All data describe the United States equity market and U.S. macro-financial indicators. The target is an anonymized 24-month U.S. window; the reference is an anonymized, widely recognized U.S. bubble prototype.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the NL summaries of the reference and target files.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A reference bubble prototypes represented as natural-language summaries for each of six indicators.

- A target 24-month window is represented the same way. This target window is based on an unknown period.

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

4) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

5) Your task (free qualitative reasoning):

- Inputs to use: the de-identified NL summaries for the Target 24-month window and the Reference bubble prototype (pattern-level, anonymized).

- Goal: Assign a single probability that the Target 24-month window is in a bubble state, using only the information provided here.

- Method: Read the provided summaries carefully and compare them for the stated goal. Do not rely on any external knowledge, numeric thresholds, or unstated calibration rules.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so explicitly and avoid confident extremes.

6) Output (exactly this format):

- Probability: <a single number in [0,1] with four decimals representing P(bubble).

- Rationale: Justify your decision based only on the provided natural summaries of the macro-financial indicators of the references.

7) Data (de-identified NL summaries):

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The bubble data has tag “Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here:

2. Non-Bubble Only

2-1. Standalone LLM Prompot (Configured with Three Non-Bubble and Fed with One Non-Bubble Data)

Act as an impartial, closed-world evaluator: using only the de-identified natural-language summaries of six U.S. macro-financial indicators provided below—without external or historical knowledge—assign a probability that the target 24-month window is NOT a bubble by strictly following the rules specified in the continuing sections.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat the reference as an anonymized, widely recognized non-bubble prototype; treat the target as an anonymized window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- All data describe the United States equity market and U.S. macro-financial indicators. The target is an anonymized 24-month U.S. window; the reference is an anonymized, widely recognized U.S. non-bubble prototype.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the NL summaries of the reference and target files.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A reference 24-month non-bubble prototypes represented as natural-language summaries for each of six indicators.

- A target 24-month window is represented the same way. This target window is based on an unknown period.

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

4) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

5) Your task (free qualitative reasoning):

- Inputs to use: the de-identified NL summaries for the Target 24-month window and the Reference non-bubble prototype (pattern-level, anonymized).

- Goal: Assign a single probability that the Target 24-month window is in a non-bubble state, using only the information provided here.

- Method: Read the provided summaries carefully and compare them for the stated goal. Do not rely on any external knowledge, numeric thresholds, or unstated calibration rules.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so explicitly and avoid confident extremes.

6) Output (exactly this format):

- Probability: <a single number in [0,1] with four decimals representing P(non-bubble).

- Rationale: Justify your decision based only on the provided natural summaries of the macro-financial indicators of the references.

7) Data (de-identified NL summaries):

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag “Non-Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here: {target}   {Bubble_Prototypes}

2-2. Standalone LLM Prompt (Configured with Four Non-Bubble and Fed with One Bubble Data)

Act as an impartial, closed-world evaluator: using only the de-identified natural-language summaries of six U.S. macro-financial indicators provided below—without external or historical knowledge—assign a probability that the target 24-month window is NOT a bubble by strictly following the rules specified in the continuing sections.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat the reference as an anonymized, widely recognized non-bubble prototype; treat the target as an anonymized window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- All data describe the United States equity market and U.S. macro-financial indicators. The target is an anonymized 24-month U.S. window; the reference is an anonymized, widely recognized U.S. non-bubble prototype.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the NL summaries of the reference and target files.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A reference 24-month non-bubble prototypes represented as natural-language summaries for each of six indicators.

- A target 24-month window is represented the same way. This target window is based on an unknown period.

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

4) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

5) Your task (free qualitative reasoning):

- Inputs to use: the de-identified NL summaries for the Target 24-month window and the Reference non-bubble prototype (pattern-level, anonymized).

- Goal: Assign a single probability that the Target 24-month window is in a non-bubble state, using only the information provided here.

- Method: Read the provided summaries carefully and compare them for the stated goal. Do not rely on any external knowledge, numeric thresholds, or unstated calibration rules.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so explicitly and avoid confident extremes.

6) Output (exactly this format):

- Probability: <a single number in [0,1] with four decimals representing P(non-bubble).

- Rationale: Justify your decision based only on the provided natural summaries of the macro-financial indicators of the references.

7) Data (de-identified NL summaries):

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag “Non-Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here: {json_payload}

3. Bubble & Non-Bubble

3-1. Standalone LLM Prompt (Configured with 4 Bubble Data & 3 Non-Bubble Data and Fed with Non-Bubble Data)

Act as an impartial, closed-world evaluator: using only the de-identified natural-language summaries of six U.S. macro-financial indicators provided below—without external or historical knowledge—assign a probability that the target 24-month window is a bubble by strictly following the rules specified in the continuing sections.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat the reference as an anonymized, widely recognized bubble and non-bubble prototype; treat the target as an anonymized window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- All data describe the United States equity market and U.S. macro-financial indicators. The target is an anonymized 24-month U.S. window; the reference is an anonymized, widely recognized U.S. bubble and non-bubble prototype.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the NL summaries of the reference and target files.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A reference bubble and non- bubble prototypes represented as natural-language summaries for each of six indicators.

- A target 24-month window is represented the same way. This target window is based on an unknown period.

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

4) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

5) Your task (free qualitative reasoning):

- Inputs to use: the de-identified NL summaries for the Target 24-month window and the Reference bubble and non-bubble prototype (pattern-level, anonymized).

- Goal: Assign a single probability that the Target 24-month window is in a bubble state, using only the information provided here.

- Method: Read the provided summaries carefully and compare them for the stated goal. Do not rely on any external knowledge, numeric thresholds, or unstated calibration rules.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so explicitly and avoid confident extremes.

6) Output (exactly this format):

- Probability: <a single number in [0,1] with four decimals representing P(bubble).

- Rationale: Justify your decision based only on the provided natural summaries of the macro-financial indicators of the references.

7) Data (de-identified NL summaries):

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag

“Non-Bubble Prototype”, while the bubble data has tag “Bubble Prototype”. Your target data for output has tag “Target Data”. It can be assessed here: {target}   {Bubble_Prototypes}

3-2. Standalone LLM Prompt (Configured with 3 Bubble Data & 4 Non-Bubble Data and Fed with Bubble Data)

Act as an impartial, closed-world evaluator: using only the de-identified natural-language summaries of six U.S. macro-financial indicators provided below—without external or historical knowledge—assign a probability that the target 24-month window is a bubble by strictly following the rules specified in the continuing sections.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat the reference as an anonymized, widely recognized bubble and non-bubble prototype; treat the target as an anonymized window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- All data describe the United States equity market and U.S. macro-financial indicators. The target is an anonymized 24-month U.S. window; the reference is an anonymized, widely recognized U.S. bubble and non-bubble prototype.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the NL summaries of the reference and target files.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A reference bubble and non-bubble prototypes represented as natural-language summaries for each of six indicators.

- A target 24-month window is represented the same way. This target window is based on an unknown period.

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

4) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

5) Your task (free qualitative reasoning):

- Inputs to use: the de-identified NL summaries for the Target 24-month window and the Reference bubble and non-bubble prototype (pattern-level, anonymized).

- Goal: Assign a single probability that the Target 24-month window is in a bubble state, using only the information provided here.

- Method: Read the provided summaries carefully and compare them for the stated goal. Do not rely on any external knowledge, numeric thresholds, or unstated calibration rules.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so explicitly and avoid confident extremes.

6) Output (exactly this format):

- Probability: <a single number in [0,1] with four decimals representing P(bubble).

- Rationale: Justify your decision based only on the provided natural summaries of the macro-financial indicators of the references.

7) Data (de-identified NL summaries):

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag “Non-Bubble Prototype”, while the bubble data has tag “Bubble Prototype”. Your target data for output has tag “Target Data”. It can be assessed here: {target}   {Bubble_Prototypes}

2. DL-LLM Architecture Prompts

2. Based on Transformer DL

1-1. DL - LLM Prompt (Bubble-Only Configuration Fed with Bubble Data Output)

Act as an impartial, closed-world evaluator: using only (i) the DL model’s bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any “reference bubble prototype” as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability $p_{DL} \in [0,1]$ produced by a DL model trained on multiple historical bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, representing a bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar bubble probability $p \in [0,1]$ for that window. The model is trained on multiple historical bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Inputs are standardized 24×6 windows over CPI, PPI, FEDFUNDS, DGS10, SP500_PE, DJIA. A 2-layer bidirectional LSTM with hidden size 128 per direction encodes each window; the final forward/backward states are concatenated (256) and passed through a Linear(256→128), then $\ell 2$-normalized to produce the embedding z. A classifier MLP (128→64→32→1, ReLU, dropout 0.2) outputs $p_{bubble} \in [0,1]$ via Sigmoid. Training draws windows that do not cross file (“prototype”) boundaries; for each window it forms an augmented view using TimeWarp + Drift + AddNoise. The objective is NT-Xent contrastive loss (τ=0.05) between the two views + 0.5× BCE with label 1 (bubble) on both views. Optimization uses Adam (lr=3e-4) for 300 epochs, batch size ≤ 64. Feature scaling uses StandardScaler (macro vs. market columns fit on train data). At inference, the model reports pbubble for each sliding window (typically the last 24-month window or a summary such as the mean).

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The bubble data has tag “Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here:

{target}

{Bubble_Prototypes}

1-2. DL - LLM Prompt (Bubble-Only Configuration and Fed with Non-Bubble Data Output)

Act as an impartial, closed-world evaluator: using only (i) the DL model’s bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target

24-month window is a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference bubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, representing a bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar bubble probability p∈[0,1] for that window. The model is trained on multiple historical bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Each sample is a standardized 24×6 window over CPI, PPI, FEDFUNDS, DGS10, SP500_PE, DJIA. A 2-layer bidirectional LSTM (hidden=128 per direction) encodes the sequence; the final forward/backward states are concatenated (256), projected with Linear(256→128), then ℓ2-normalized to an embedding. A classifier head (MLP 128→64→32→1, ReLU, dropout 0.2) outputs pbubble∈[0,1] via Sigmoid. Training forms windows that do not cross file ("prototype") boundaries, creates an augmented view with TimeWarp + Drift + AddNoise, and optimizes NT-Xent contrastive loss (τ=0.05) + 0.5× BCE with label 1 (bubble) for both views. Optimization: Adam (lr=3e-4), 300 epochs, batch size ≤ 64; features are scaled with StandardScaler (macro vs. market columns fit on train data). At inference, the model returns pbubble for each sliding window; for reporting, use the last 24-month score (or an aggregate such as the mean across windows).

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., "+7.4") or "uncertain" when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The bubble data has tag “Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here:
{target}   {Bubble_Prototypes}

2-1. DL - LLM Prompt (Configured with Three Non-Bubble and Fed with One Non-Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model's non-bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is NOT a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference Non-bubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical non-bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference Non-bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, representing a non-bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar non-bubble probability p∈[0,1] for that window. The model is trained on multiple historical non-bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Each sample is a standardized 24×6 window over CPI, PPI, FEDFUNDS, DGS10, DJIA, SP500_PE. A 2-layer bidirectional LSTM (128 hidden/dir) encodes the sequence; the last forward/backward states are concatenated (256), passed through Linear(256→128), then ℓ2-normalized to an embedding z. A classifier head (MLP 128→64→32→1, ReLU, dropout 0.2) produces pbubble∈[0,1] via Sigmoid. Training forms windows that do not cross prototype (file) boundaries, creates an augmented view with TimeWarp + Drift + AddNoise, and optimizes NT-Xent contrastive loss (τ=0.05) + 0.5× BCE with label 1 (bubble) for both views. Optimization: Adam (lr=3e-4), 300 epochs, batch ≤ 64. Features are scaled with StandardScaler (fit on training only, macro and market columns separately). Inference: compute pbubblep_{\text{bubble}}pbubble for each sliding 24-month window in the evaluation CSV; report the latest-window score (or an aggregate such as mean/max) as the non-bubble probability.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Non-Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(non-bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag “Non-Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here: {target} {Bubble_Prototypes}

2-2. DL-LLM Prompt (Configured with Four Non-Bubble and Fed with Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model's non-bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is NOT a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference nonbubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical non-bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference non-bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, representing a non-bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar non-bubble probability $p \in [0,1]$ for that window. The model is trained on multiple historical non-bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Each sample is a standardized 24×6 window over CPI, PPI, FEDFUNDS, DGS10, DJIA, SP500_PE. A 2-layer bidirectional LSTM (128 hidden/dir) encodes the sequence; the last forward/backward states are concatenated (256), projected with Linear(256→128), then ℓ2-normalized to an embedding z. A classifier head (MLP 128→64→32→1, ReLU, dropout 0.2) outputs pnon-bubble∈[0,1] via Sigmoid. Windows do not cross prototype (file) boundaries; an augmented view is built with TimeWarp + Drift + AddNoise. The objective is NT-Xent contrastive loss (τ=0.05) + 0.5× BCE with label 1 (non-bubble) for both views. Optimization uses Adam (lr=3e-4) for 300 epochs, batch size ≤ 64. Features are scaled with StandardScaler (fit on training only; macro and market columns scaled separately). Inference: treat pnon-bubblep_{\text{non-bubble}}pnon-bubble as a similarity to the non-bubble regime; compute it for each sliding 24-month window in the evaluation CSV and report the latest (or an aggregate such as mean/max). For the paper's bubble score, report pbubble=1−pnon-bubble.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Non-Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(non-bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag "Non-Bubble Prototype", while your target data for output has tag "Target Data". It can be assessed here: {json_payload}

3-1. DL - LLM Prompt (Configured with 4 Bubble Data & 3 Non-Bubble Data and Fed with Non-Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model's bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference bubble and non-bubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical non-bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, where some represent a bubble pattern while others represent a non-bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar bubble probability $p \in [0,1]$ for that window. The model is trained on multiple historical non-bubble and bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Each sample is a standardized 24×6 window over CPI, PPI, FEDFUNDS, DGS10, DJIA, SP500_PE. A 2-layer bidirectional LSTM (128 hidden/dir) encodes the sequence; final forward/backward states are concatenated (256), projected 256→128, then ℓ2-normalized to an embedding z. A classifier head (MLP 128→64→32→1, ReLU, dropout 0.2) outputs $p_{bubble} \in [0,1]$ via Sigmoid. Windows never cross prototype (file) boundaries; an augmented view is created with TimeWarp + Drift + AddNoise. Training minimizes NT-Xent contrastive loss (τ=0.05) + 0.5× BCE using the true labels (bubble=1, non-bubble=0) for both views. Optimization: Adam (lr=3e-4) for 300 epochs, batch size ≤ 64. Features are standardized with StandardScaler (fit on training only; macro and market columns scaled separately). Inference/reporting: compute p_{bubble} for each sliding 24-month window of a target CSV and report the latest probability (or an aggregate such as mean/max) per your evaluation protocol.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., "+7.4") or "uncertain" when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Bubble and Non-Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tasg "Non-Bubble Prototype", while the bubble data has tag "Bubble Prototype". Your target data for output has tag "Target Data". It can be assessed here: {target}   {Bubble_Prototypes}

3-2. DL - LLM Prompt (Configured with 3 Bubble Data & 4 Non-Bubble Data and Fed with Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model's bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference bubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical non-bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, where some represent a bubble pattern while others represent a non-bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar non-bubble probability $p \in [0,1]$ for that window. The model is trained on multiple historical non-bubble and bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Each example is a standardized 24×6 window over CPI, PPI, FEDFUNDS, DGS10, DJIA, SP500_PE. A 2-layer bidirectional LSTM (hidden=128/dir) encodes the window; the final forward/backward states are concatenated (256), projected 256→128, then ℓ2-normalized to an embedding z. A classifier head (MLP 128→64→32→1, ReLU, dropout 0.2) outputs $p_{bubble} \in [0,1]$ (Sigmoid). Windows never cross prototype/file boundaries; a second "view" is formed with TimeWarp + Drift + AddNoise. Training minimizes NT-Xent (τ=0.05) between the two views + 0.5× BCE with the true labels (non-bubble=0, bubble=1) for both views. Optimization: Adam (lr=3e-4) for 300 epochs, batch size ≤ 64. Features are standardized with StandardScaler (fit on training only; macro and market columns scaled separately). Inference/reporting: compute pbubble for each sliding 24-month window of a target CSV and report the latest probability (or an aggregate such as mean/max) per your evaluation protocol.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Bubble and Non-Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag “Non-Bubble Prototype”, while the bubble data has tag “Bubble Prototype”. Your target data for output has tag “Target Data”. It can be assessed here: {target}   {Bubble_Prototypes}

2. Based on Transformer DL

1. Bubble Only

1-1. DL - LLM Prompt (Configured with Bubble-Only Data and Fed with Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model’s bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any “reference bubble prototype” as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, representing a bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar bubble probability p∈[0,1] for that window. The model is trained on multiple historical bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Inputs are standardized 24×6 windows over CPI, PPI, FEDFUNDS, DGS10, SP500_PE, DJIA. The encoder is a Transformer: an input projection Linear(6→128) feeds a 2-layer TransformerEncoder (nhead = 4, dropout = 0.1, norm_first = True) with sinusoidal positional embeddings; pooling is configurable and set to "last" by default (alternatives "mean"/"cls"; the latter prepends a learnable [CLS] token). The resulting 128-d embedding z is ℓ2 normalized and passed to a classifier MLP (128→64→32→1 with ReLU and dropout 0.2) whose Sigmoid outputs pbubble. Each step encodes both the anchor and its augmented view, computes NT-Xent on the embeddings, and applies BCE on each view's probability against the bubble = 1 label; the total loss is L=LNT-Xent+0.5(BCEanchor+BCEaug). Seeds are fixed; device selection is automatic (CUDA if available). The saved package includes model weights, encoder configuration (window length, pooling, Transformer hyperparameters), and fitted scalers, enabling consistent inference of pbubble on future 24-month windows.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The bubble data has tag “Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here:
{target}   {Bubble_Prototypes}

1-2. DL - LLM Prompt (Configured with Bubble-Only Data and Fed with Non-Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model’s bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference bubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, representing a bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar bubble probability p∈[0,1] for that window. The model is trained on multiple historical bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Architecture for the DL Model Training: Inputs are standardized 24×6 windows over CPI, PPI, FEDFUNDS, DGS10, SP500_PE, DJIA. The encoder is a Transformer: an input projection Linear(6→128) feeds a 2-layer TransformerEncoder (nhead = 4, dropout = 0.1, norm_first = True) with sinusoidal positional embeddings; pooling is configurable and set to "last" by default (alternatives "mean"/"cls"; the latter adds a learnable [CLS] token). The resulting 128-d embedding z is ℓ2-normalized and passed to a classifier MLP (128→64→32→1 with ReLU and dropout 0.2) whose Sigmoid outputs pbubble. Each step encodes both anchor and augmented sequences, computes NT-Xent on their embeddings, and applies BCE on each view's probability against the fixed bubble = 1 label; the total loss is L=LNT-Xent+0.5(BCEanchor+BCEaug). Seeds are fixed; device selection is automatic (CUDA if available). The saved package includes model weights, encoder configuration (window length, pooling, Transformer hyperparameters), fitted scalers, required column names, and the list of training prototypes, enabling consistent inference of pbubble on future 24-month windows.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The bubble data has tag “Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here:
{target}   {Bubble_Prototypes}

2. Non-Bubble Only

2-1. DL - LLM Prompt (Configured with Non-Bubble-Only Data and Fed with Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model's non-bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a non-bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference non-bubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical non-bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference non-bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, representing a non-bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar non-bubble probability $p \in [0,1]$ for that window. The model is trained on multiple historical non-bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Inputs are standardized 24×6 windows over CPI, PPI, FEDFUNDS, DGS10, SP500_PE, DJIA. The encoder is a Transformer: an input projection Linear(6→128) feeds a TransformerEncoder with 2 encoder layers (nhead = 4, dropout = 0.1, norm_first = True) and sinusoidal positional embeddings; pooling is configurable and set to "last" by default (alternatives "mean"/"cls", where "cls" prepends a learnable token). The resulting 128-d embedding z is ℓ2-normalized and passed to a classifier MLP (128→64→32→1 with ReLU and dropout 0.2), whose Sigmoid outputs pnon. Each step encodes both anchor and augmented sequences, computes NT-Xent on their embeddings, and applies BCE on each view's probability against the fixed non-bubble = 1 label; the total loss is L=LNT-Xent+0.5(BCEanchor+BCEaug). Seeds are fixed; device selection is automatic (CUDA if available). The saved package includes weights, encoder configuration (window length, pooling, Transformer hyperparameters), fitted scalers, required column names, and the list of training prototypes, enabling consistent inference where pbubble=1−pnonp on future 24-month windows.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Non-Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(non-bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag "Non-Bubble Prototype", while your target data for output has tag "Target Data". It can be assessed here: {target} {Bubble_Prototypes}

2-2. DL-LLM Prompt (Configured with Non-Bubble-Only Data and Fed with Non-Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model's non-bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a non-bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference non-bubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability $pDL \in [0,1]$ produced by a DL model trained on multiple historical non-bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference non-bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, representing a non-bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar non-bubble probability $p \in [0,1]$ for that window. The model is trained on multiple historical non-bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Inputs are standardized 24×6 windows over CPI, PPI, FEDFUNDS, DGS10, SP500_PE, DJIA. The encoder is a Transformer: an input projection Linear(6→128) feeds a 2-layer TransformerEncoder (nhead = 4, dropout = 0.1, norm_first = True) with sinusoidal positional embeddings; pooling is configurable and set to “last” by default (alternatives “mean”/“cls”; the latter adds a learnable [CLS] token). The resulting 128-d embedding z is ℓ2-normalized and passed to a classifier MLP (128→64→32→1 with ReLU and dropout 0.2) whose Sigmoid outputs pnon. Each training step encodes both the anchor and its augmented view, computes NT-Xent on the pair of embeddings, and applies BCE on each view against the non-bubble (1) label; the total loss is L=LNT-Xent+0.5(BCEanchor+BCEaug). Seeds are fixed; device selection is automatic (CUDA if available). The saved package includes model weights, encoder config (window length, pooling, Transformer hyperparameters), fitted scalers, required column names, and the list of training prototypes, enabling consistent inference where pbubble=1−pnonp on future 24-month windows.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Non-Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(non-bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag “Non-Bubble Prototype”, while your target data for output has tag “Target Data”. It can be assessed here: {target}   {Bubble_Prototypes}

3. Bubble & Non-Bubble

3-1. DL - LLM Prompt (Configured with 4 Bubble Data & 3 Non-Bubble Data and Fed with Non-Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model’s bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any “reference bubble and non-bubble prototype” as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical bubble and non-bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference bubble and non-bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, where some represent a bubble pattern while others represent a non-bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar bubble probability p∈[0,1] for that window. The model is trained on **all** seven historical datasets (four from bubble periods and three from non-bubble periods) to learn their distinct temporal patterns. The training process uses both the original data and an augmented view. The model is trained to minimize a combined loss function of NT-Xent contrastive loss and a weighted Binary Cross-Entropy (BCE) loss using the true labels.

- Architecture for the DL Model Training: Architecture for the DL Model Training: Inputs are standardized 24×6 windows over CPI, PPI, FEDFUNDS, DGS10, SP500_PE, DJIA. The encoder used here is a Transformer: an input projection Linear(6→128) feeds a 2-layer TransformerEncoder (nhead = 4, dropout = 0.1, norm_first = True) with sinusoidal positional embeddings; pooling is configurable and set to “last” by default (alternatives “mean”/“cls” exist, with a learnable [CLS] token only used if “cls” is selected). The resulting 128-d embedding z is ℓ2-normalized and passed to a classifier MLP (128→64→32→1 with ReLU and dropout 0.2) whose Sigmoid outputs pbubble. During each step, the model encodes both the anchor and its augmented view, computes NT-Xent on the pair of embeddings, and applies BCE on each view’s probability against the inherited label; the total loss is L=LNT-Xent+0.5(BCEanchor+BCEaug). Seeds are fixed; training/device selection is automatic (CUDA if available). After training, a package is saved containing the model weights, encoder configuration (including window length, pooling, and Transformer hyperparameters), and the fitted scalers, enabling consistent inference of pbubble on future 24-month windows. (All architectural specifics are derived from the provided code.)

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., "+7.4") or "uncertain" when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Bubble and Non-Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag "Non-Bubble Prototype", while the bubble data has tag "Bubble Prototype". Your target data for output has tag "Target Data". It can be assessed here: {target}   {Bubble_Prototypes}

3-2. DL - LLM 프롬프트 (Configured with 3 Bubble Data & 4 Non-Bubble Data and Fed with Bubble Data)

Act as an impartial, closed-world evaluator: using only (i) the DL model's bubble score computed from a standardized 24-month window and (ii) the de-identified natural-language summaries for that same window plus other reference data, assign a probability that the target 24-month window is a bubble, strictly following the rules below and without any external or historical knowledge.

1) Operational constraints (read carefully):

- Closed world. Use only the information provided in this prompt. Do not use external tools, web search, prior knowledge of specific events, or any information not explicitly included here.

- No historical inference. Do not name or infer specific eras, crises, tickers, or dates. Treat any "reference bubble and non-bubble prototype" as an anonymized, pattern-level abstraction; treat the target strictly as an anonymized evaluation window.

- Perform any internal reasoning privately; output only the requested fields. Solely rely on your reasoning capabilities.

- Do not invent numbers or facts. In the rationale, where you should justify your final decision, base your explanation on the architecture and information about the DL model.

- If you regard the given reference and target evidence are weak or mixed, avoid confident extremes.

2) What you will be given (all de-identified):

- A single probability pDL∈[0,1] produced by a DL model trained on multiple historical bubble and non-bubble episodes. This score pertains only to the target 24-month window.

- Target window (24 months) summaries: De-identified natural-language summaries for the six macro-financial indicators computed over one contiguous 24-month U.S. window.

- Reference bubble and non-bubble prototype summaries: De-identified natural-language summaries of the same six indicators and feature schema, where some represent a bubble pattern while others represent a non-bubble pattern (learned at training time).

- Indicators (six) in those reference and target windows: Consumer Price Index (CPI), Producer Price Index (PPI), Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E Ratio (SP500_PE), Dow Jones Industrial Average (DJIA).

3) About the DL Model and Its Data

- What the DL model does (generic): A supervised time-series model that ingests a standardized 24-month window of six macro-financial indicators—CPI, PPI, Federal Funds Rate (FEDFUNDS), 10-Year Treasury Yield (DGS10), S&P 500 P/E (SP500_PE), and DJIA—and outputs a scalar non-bubble probability p∈[0,1] for that window. The model is trained on multiple historical non-bubble episodes to learn their common temporal patterns and validated on a held-out bubble episode to assess out-of-sample generalization; for reporting, we use the probability from the last 24 months of the evaluation period.

- Architecture for the DL Model Training: Inputs are standardized 24×6 windows over CPI, PPI, FEDFUNDS, DGS10, SP500_PE, DJIA. The encoder is a Transformer: an input projection Linear(6→128) followed by a 2-layer TransformerEncoder (nhead = 4, dropout = 0.1, norm_first = True) with sinusoidal positional embeddings; pooling is configurable and set to “last” by default (alternatives “mean”/“cls”; the latter adds a learnable [CLS] token). The resulting 128-d embedding z is ℓ2-normalized and fed to a classifier MLP (128→64→32→1 with ReLU and dropout 0.2) whose Sigmoid outputs pbubblep_{\text{bubble}}pbubble. Each training step encodes both the anchor and its augmented view, computes NT-Xent on the pair of embeddings, and applies BCE on each view’s probability against the inherited label; the total loss is L=LNT-Xent+0.5(BCEanchor+BCEaug). Seeds are fixed; device selection is automatic (CUDA if available). After training, a package is saved with model weights, encoder config (window length, pooling, Transformer hyperparameters), fitted scalers, required column names, and the list of training prototypes, enabling consistent inference of pbubble on future 24-month windows.

4) Example Natural-language summary format (per indicator, already computed for you):

- net_change_pct: The total percentage change over the full series (e.g., +18.2).

- slope_ols_pct_per_mo: OLS slope of monthly % change (e.g., +0.75).

- trend_r2: R^2 of the overall trend (e.g., 0.61).

- up_month_share: The fraction of up months (e.g., 0.63).

- vol_std_pct: The standard deviation of monthly % changes (e.g., 4.2).

- vol_late_minus_early_pct: The standard deviation of the late period minus the standard deviation of the early period, in % (e.g., +1.1).

- max_drawup_pct: The largest peak-to-prior-trough rise within the full series, % (e.g., +22.0).

- t_peak: The month index of that max drawup in the full series [1..N] (e.g., 19).

- max_drawdown_pct: The largest trough-from-prior-peak fall within the full series, % (e.g., −9.0).

- t_trough: The month index of that max drawdown in the full series [1..N] (e.g., 7).

- acf1: The lag-1 autocorrelation of monthly % changes (e.g., 0.31).

5) Guidance for the NL summaries (attached to this prompt)

• Format: The Data block contains an anonymized 24-month target window. It also includes the full reference data. Both the target and reference data are encoded as an indicator-to-feature pair. No specific filenames or dates are included. Filenames are designed with a specific convention to assist you in identifying whether a file is a target or a reference.

• Values: Each feature is a formatted string (e.g., “+7.4”) or “uncertain” when it cannot be computed without imputing missing values.

• Missing data: If present, missing_data lists month indices in [1..N] where a level is missing. This happens when a particular indicator was not officially published for that month. Do not infer across gaps. Drawup/drawdown is computed only within contiguous valid segments.

6) Your task (free qualitative and rigorous reasoning):

- Inputs to use: the DL model output for the Target 24-month window, plus de-identified NL summaries for the Target (24 months) and the Reference Bubble and Non-Bubble Prototype (pattern-level, anonymized).

- Goal: Decide whether the Target 24-month window is more consistent with a bubble or a non-bubble state, using only the provided inputs.

- Method (qualitative): Read the Target summaries and the Reference prototype summaries and compare their overall patterns. Do not use any external knowledge, numeric thresholds, or unstated calibration rules.

- Use of DL output: Treat the DL output as one piece of evidence. If the textual comparison supports or contradicts it, state that plainly in your rationale.

- Uncertainty: If the evidence is weak, mixed, or marked uncertain, say so and avoid confident extremes. Do not invent numbers or facts, and do not infer dates, eras, or tickers.

7) Output (exactly this format):

- Probability: a single number in 0,1 with four decimals representing P(bubble) for the Target 24-month window

- Rationale: Justify your decision based only on the provided DL model output and natural summaries of the macro-financial indicators of the references, noting whether that quoted evidence supports or contradicts the DL output.

8) Data (DL Model Output and de-identified NL summaries):

The DL Model outputted:

The de-identified NL summaries are given as a JSONL format with specific tags attached that will allow you to distinguish the reference and target test files. The non-bubble data has tag “Non-Bubble Prototype”, while the bubble data has tag “Bubble Prototype”. Your target data for output has tag “Target Data”. It can be assessed here: {target}   {Bubble_Prototypes}