

Recommended analytical logic

This workflow treats the microblog stream as an **early-warning signal**, not as confirmed clinical case data. It therefore combines automated extraction with repeated visual and epidemiological review.

1. **Validate the illness signal.** Extract messages likely describing current flu-like illness, while distinguishing first-person symptom reports from news, jokes, retweets, or general discussion. Analysts inspect mapped and temporal samples, then refine the detection criteria until the signal is credible.
2. **Create population-normalised space-time outbreak units.** Aggregate relevant posts into zone-by-time episodes (for example, 6- or 12-hour bins), compute counts and rates per population, and derive onset time, growth rate, local spread, and contextual features such as distance to water, hospitals, landmarks, and highways.
3. **Identify plausible origin areas.** Rank candidate source zone-time locations using sustained early excess, subsequent outward propagation, and consistency with the observed timeline. Analysts inspect maps, timelines, and early-message text before accepting an origin hypothesis.
4. **Compare competing transmission explanations.** Fit and visually compare candidate spread models:
 - **Person-to-person / local-contact:** propagation through adjacent, dense zones or along transport corridors.
 - **Airborne:** downwind, time-lagged propagation aligned with wind direction and speed.
 - **Waterborne:** concentration near, or progression along, water-connected areas.
 - **Other / reporting artefact:** common-source exposure near landmarks, hospital-related posting bias, or another unobserved mechanism.

The team evaluates each mechanism by its fit, spatial-temporal residuals, and epidemiological plausibility. It iteratively adjusts temporal resolution, lag ranges, thresholds, and covariates where needed.

5. **Produce conditional operational guidance.** The final output ranks likely origin areas and mechanisms, states uncertainty explicitly, and links each hypothesis to immediate action: targeted case investigation, mobile clinical support, water sampling, air-quality/environmental investigation, risk communication, or surveillance expansion.

This design adapts the spatial clustering and aggregation strategy of “**Extracting significant places from trajectories**”, the iterative model-and-residual assessment pattern of “**Spatio-temporal analysis and modelling**”, and the lag/sensitivity calibration

logic of “**Detecting Stable Cross-Impact Patterns in Bivariate Time Series.**” The coordinated map–timeline–space-time exploration is also informed by “**ISeeCube: Visual Analysis of Gaze Data for Video.**”

Crucially, the available data can support **evidence-ranked hypotheses**, but cannot prove a transmission route. Confirmatory epidemiological investigation, clinical case data, environmental samples, and contact-tracing information remain necessary.

ATWL representation

atwl

```
workflow vastopolis-outbreak-origin-and-transmission-analysis
  template: characterise
    loop(visualise abstract assess
      generate-knowledge characterise)
    define-unit
    loop(define-unit characterise contextualise visualise
      build-model characterise visualise abstract
      assess generate-knowledge)
    generate-knowledge

description: "Human-in-the-loop visual analytics workflow for detecting
  a flu-like illness outbreak from geolocated microblogs, estimating
  plausible origin areas, comparing transmission-mechanism hypotheses,
  and producing uncertainty-aware emergency-response guidance."

# =====
# INPUT
# =====

artifact D_posts : entities
  origin: given
  internal structure: elementary
  embedment: {set, space, time}
  features:
    - id: timestamp
      value structure: atomic
      value type: temporal
      description: "Posting time"
    - id: location
      value structure: atomic
```

```
    value type: spatial
    description: "GPS coordinate recorded by the mobile device"
  - id: text
    value structure: atomic
    value type: text
    description: "Microblog message content"
description: "Geolocated and timestamped microblog messages posted in
Vastopolis during the outbreak observation period."
```

artifact D_zones : entities

```
origin: given
internal structure: region
embedment: space
features:
  - id: zone_name
    value structure: atomic
    value type: categorical
    description: "Identifier of one of the 13 city zones"
  - id: population
    value structure: atomic
    value type: numeric
    description: "Resident population of the zone"
description: "Metropolitan analysis zones with population statistics."
```

artifact D_map_features : entities

```
origin: given
internal structure: elementary
embedment: space
features:
  - id: feature_type
    value structure: atomic
    value type: categorical
    description: "Highway, hospital, landmark, water body, or other ma
  - id: geometry
    value structure: atomic
    value type: spatial
    description: "Spatial geometry of the map feature"
description: "Mapped highways, hospitals, landmarks, and water bodies."
```

artifact D_weather : entities

```
origin: given
internal structure: elementary
```

```
embedment: time
features:
  - id: wind_speed
    value structure: atomic
    value type: numeric
    description: "Observed wind speed"
  - id: wind_direction
    value structure: atomic
    value type: numeric
    description: "Observed wind direction"
description: "Daily weather observations for Vastopolis."
```

```
artifact K_epidemiology : knowledge(D_posts, D_zones, D_weather)
origin: given
representation form: "domain assumptions and interpretation criteria"
description: "Public-health knowledge concerning flu-like symptoms,
  plausible incubation and reporting delays, expected spread mechanism
  intervention constraints, and known limitations of social-media sign
```

```
artifact S_signal_detection : specification
origin: given
representation form: "text-classification and filtering settings"
description: "Initial syndrome vocabulary, symptom phrases, language
  rules, confidence threshold, and exclusion rules for news, metaphori
  historical, or non-local illness references."
```

```
artifact S_outbreak_analysis : specification
origin: given
representation form: "spatio-temporal and mechanism-model settings"
description: "Initial time-bin size, outbreak threshold, smoothing met
  candidate lag range, spatial neighbourhood definition, and assumptio
  for person-to-person, airborne, waterborne, and other-source models.
```

```
# =====
# STEP 1: INITIAL SYNDROME-SIGNAL EXTRACTION
# =====
```

```
transform T_extract_signal :
  intent: characterise
  manner: "text classification and symptom extraction"
  input: D_posts, S_signal_detection, K_epidemiology
  output: F_illness_signal
```

```

actor: hybrid
description: "Score each message for evidence of a contemporaneous,
    first-person or locally observed flu-like illness report; retain
    symptom terms, confidence, and exclusion rationale."

artifact F_illness_signal : feature(D_posts)
    value structure: vector
    value type: {numeric, categorical, text}
    description: "Per-message illness relevance score, symptom indicators,
        report category, and exclusion reason."

# =====
# STEP 2: SIGNAL-QUALITY REFINEMENT LOOP
# =====

loop L_signal_refinement:
    purpose: "Ensure that microblog-derived illness reports provide a credible
        outbreak signal rather than primarily news, noise, or reporting artefacts"
    until: "The analyst judges that retained messages plausibly indicate
        current local illness and that obvious false-positive patterns are corrected"
    body:

        transform T_visualise_signal :
            intent: visualise
            manner: "linked message map, timeline, and text inspection"
            input: D_posts, F_illness_signal, D_zones
            output: V_signal
            actor: machine
            description: "Show illness-score distributions over time and space
                representative message text, and linked zone summaries."

        artifact V_signal : visualisation(D_posts, F_illness_signal, D_zones)
            layout: "geographic map, temporal histogram, ranked message list"
            form: "points, bars, and text records"
            encoding: "map position from GPS coordinates; point colour from
                illness confidence; temporal position from timestamp; text panel
                shows message content and classification rationale"
            description: "Coordinated view for inspecting whether extracted
                messages plausibly reflect an emerging outbreak."

        transform T_interpret_signal :
            intent: abstract

```

```

manner: "perception-and-interpretation of message validity"
input: V_signal, K_epidemiology
output: P_signal_characteristics
actor: human
description: "Identify false-positive themes, reporting biases,
    plausible symptom language, and spatial or temporal concentration

artifact P_signal_characteristics : pattern(D_posts)
    representation form: "descriptions of reliable and unreliable messages
    description: "Observed message patterns indicating credible illness
        reports, likely news or metaphor, and possible reporting biases.

transform T_assess_signal :
    intent: assess
    manner: "evaluate syndrome-detection adequacy"
    input: V_signal, P_signal_characteristics, K_epidemiology
    output: K_signal_quality
    actor: human
    description: "Judge whether the extraction settings produce a
        sufficiently credible illness signal for outbreak analysis."

artifact K_signal_quality : knowledge(F_illness_signal)
    representation form: "quality judgment"
    description: "Assessment of signal validity, noise, likely reporting
        artefacts, and need for further detection refinement."

if K_signal_quality indicates refinement needed:
    then:

        transform T_refine_signal_detection :
            intent: generate-knowledge
            manner: "revise classification and filtering criteria"
            input: K_signal_quality, P_signal_characteristics,
                S_signal_detection
            output: S_signal_detection_r
            actor: human
            description: "Adjust symptom vocabulary, confidence thresholds
                exclusion patterns, or language rules."

        artifact S_signal_detection_r : specification
            representation form: "revised text-classification settings"
            description: "Updated syndrome-detection criteria."

```

```

transform T_reextract_signal :
  intent: characterise
  manner: "reclassify messages with revised settings"
  input: D_posts, S_signal_detection_r, K_epidemiology
  output: F_illness_signal_r
  actor: machine
  description: "Recompute illness-report relevance and categorie

artifact F_illness_signal_r : feature(D_posts)
  value structure: vector
  value type: {numeric, categorical, text}
  description: "Updated illness-signal classification results."

assign:
  S_signal_detection := S_signal_detection_r
  F_illness_signal := F_illness_signal_r

else:
  exit loop L_signal_refinement

end loop L_signal_refinement

# =====
# STEP 3: DEFINE ILLNESS-REPORT ANALYSIS UNITS
# =====

transform T_select_illness_posts :
  intent: define-unit
  manner: "filter high-confidence illness reports"
  input: D_posts, F_illness_signal, S_signal_detection
  output: D_illness_posts
  actor: machine
  description: "Retain posts meeting the validated illness-report criter

artifact D_illness_posts : entities
  internal structure: elementary
  embedment: {set, space, time}
  features:
    - id: illness_confidence
      value structure: atomic
      value type: numeric

```

```

        description: "Validated confidence that the post reports current i
- id: symptom_terms
  value structure: list
  value type: text
  description: "Extracted symptom terms"
description: "Filtered geolocated microblog reports treated as an
  outbreak-surveillance signal rather than confirmed clinical cases."

# =====
# STEP 4: OUTBREAK HYPOTHESIS REFINEMENT LOOP
# =====

loop L_outbreak_hypothesis:
  purpose: "Estimate outbreak origin and compare transmission explanatio
    through iterative spatial-temporal modelling and visual residual ana
  until: "A source-area hypothesis and transmission explanation are
    sufficiently supported for provisional emergency action, with uncert
    and alternative explanations explicitly documented"
  body:

    # -----
    # 4a. Construct population-normalised zone-time episodes
    # -----

    transform T_aggregate_zone_time :
      intent: define-unit
      manner: "spatio-temporal aggregation into zone-time episodes"
      input: D_illness_posts, D_zones, S_outbreak_analysis
      output: D_zone_time, F_burden
      actor: machine
      description: "Aggregate illness reports into configurable time bin
        within zones; compute population-normalised illness-report burde

    artifact D_zone_time : entities
      internal structure: episode
      embedment: {space, time}
      features:
        - id: zone_reference
          value structure: atomic
          value type: reference
          description: "Associated city zone"
        - id: temporal_extent

```

```

        value structure: atomic
        value type: temporal
        description: "Start and end of the aggregation interval"
description: "Zone-by-time outbreak-surveillance episodes."

artifact F_burden : feature(D_zone_time)
    value structure: vector
    value type: numeric
    representation form: "counts and population-normalised rates"
    description: "Illness-post count, rate per resident population,
        symptom composition, and confidence-weighted burden per zone-time"

# -----
# 4b. Derive propagation and contextual features
# -----

transform T_characterise_spread :
    intent: characterise
    manner: "compute onset, growth, neighbourhood, and exposure features"
    input: D_zone_time, F_burden, D_map_features, D_weather,
        S_outbreak_analysis
    output: F_spread, F_context
    actor: machine
    description: "Compute sustained-onset time, growth rate, local
        propagation relations, zone adjacency, wind-aligned exposure,
        proximity to water, hospitals, landmarks, and highways."

artifact F_spread : feature(D_zone_time)
    value structure: vector
    value type: {numeric, temporal}
    description: "Temporal outbreak descriptors including baseline-adjusted
        burden, onset time, growth rate, persistence, and lagged spread
        relationships to neighbouring zones."

artifact F_context : feature(D_zone_time)
    value structure: vector
    value type: {numeric, spatial, categorical}
    description: "Contextual descriptors including population, zone
        adjacency, distance to water, highways, landmarks and hospitals,
        plus wind direction and speed during plausible exposure windows.

transform T_arrange_zone_time_map :
```

```

intent: contextualise
manner: "map-based arrangement with temporal small multiples"
input: D_zone_time, D_zones
output: A_zone_time_map
actor: machine
description: "Place each zone-time episode in its geographic zone,
    retaining time as an animation, small-multiple, or timeline dime

artifact A_zone_time_map : arrangement(D_zone_time)
    context: D_zones
    principle: "geographic zone placement with ordered temporal bins"
    description: "Map arrangement for comparing burden, onset, and gro
        across zones through time."

transform T_visualise_outbreak :
    intent: visualise
    manner: "coordinated outbreak map, timeline, and space-time view"
    input: A_zone_time_map, F_burden, F_spread, F_context,
        D_map_features, D_weather
    output: V_outbreak
    actor: machine
    description: "Visualise population-normalised burden, onset order,
        growth, contextual infrastructure, water bodies, and wind vector

artifact V_outbreak : visualisation(A_zone_time_map, F_burden, F_spr
    F_context, D_map_features, D_weather)
    layout: "linked map, zone timelines, onset-ranking chart, and
        space-time display"
    form: "coloured regions, proportional marks, lines, arrows, and ba
    encoding: "map position from zones; colour from burden or onset;
        mark size from growth; arrows from wind direction and speed;
        overlays identify water, highways, hospitals, and landmarks"
    description: "Interactive overview for visually identifying early
        clusters, propagation fronts, and contextual correspondences."

# -----
# 4c. Build source and competing-mechanism models
# -----

transform T_build_hypothesis_models :
    intent: build-model
    manner: "fit source inference and competing spread models"

```

```

input: D_zone_time, F_burden, F_spread, F_context,
      K_epidemiology, S_outbreak_analysis
output: M_origin, M_transmission
actor: hybrid
description: "Build a source-ranking model and alternative
             spatio-temporal transmission models using epidemiologically
             plausible lags and mechanisms."

artifact M_origin : model(D_zone_time, F_burden, F_spread)
  model type: "spatio-temporal source inference model"
  representation form: "candidate source zone-time likelihood surface"
  description: "Ranks candidate origin locations and times according
               to early sustained burden and subsequent outward propagation."

artifact M_transmission : model(F_spread, F_context, D_weather)
  model type: "competing transmission-mechanism models"
  representation form: "mechanism-specific lagged spatial propagation"
  description: "Alternative explanatory models for person-to-person
               local-contact spread, wind-aligned airborne spread, water-associated
               spread, and other/common-source or reporting-bias explanations."

transform T_score_hypotheses :
  intent: characterise
  manner: "estimate fit, prediction, and residual patterns"
  input: M_origin, M_transmission, D_zone_time, F_burden,
        F_spread, F_context
  output: F_origin_scores, F_mechanism_fit, F_residuals
  actor: machine
  description: "Compute candidate-origin scores, mechanism-specific
               fit statistics, predicted propagation patterns, and unexplained
               spatial-temporal residuals."

artifact F_origin_scores : feature(D_zone_time)
  value structure: atomic
  value type: numeric
  description: "Relative support for each zone-time unit as a plausible
               outbreak origin."

artifact F_mechanism_fit : feature(M_transmission)
  value structure: vector
  value type: numeric
  description: "Comparative fit, lag consistency, and uncertainty for"

```

person-to-person, airborne, waterborne, and other-source hypothe

```
artifact F_residuals : feature(D_zone_time)
  value structure: vector
  value type: numeric
  description: "Differences between observed and model-predicted bur
    or propagation, used to detect systematic unexplained patterns."
```

```
transform T_visualise_hypotheses :
  intent: visualise
  manner: "comparative origin and mechanism evidence dashboard"
  input: V_outbreak, F_origin_scores, F_mechanism_fit, F_residuals,
        M_transmission
  output: V_hypotheses
  actor: machine
  description: "Compare observed spread with mechanism-specific expe
    fronts, residual maps, source rankings, wind alignment, and lag
```

```
artifact V_hypotheses : visualisation(V_outbreak, F_origin_scores,
  F_mechanism_fit, F_residuals, M_transmission)
  layout: "coordinated source-ranking panel, observed-versus-predict
    maps, lag plots, mechanism comparison chart, and residual map"
  form: "ranked bars, maps, temporal curves, directional arrows,
    uncertainty intervals, and residual marks"
  encoding: "source rank from origin score; model colour from mechan
    observed and predicted burden shown comparatively; residual colo
    from under- or over-prediction; uncertainty shown by intervals"
  description: "Evidence dashboard for judging where the outbreak li
    started and which transmission explanation best accounts for spr
```

```
# -----
# 4d. Human interpretation and assessment
# -----
```

```
transform T_interpret_hypotheses :
  intent: abstract
  manner: "epidemiological interpretation of competing explanations"
  input: V_hypotheses, K_epidemiology
  output: P_outbreak_hypotheses
  actor: human
  description: "Interpret candidate origin areas, probable propagati
    sequence, support for each mechanism, contradictory evidence, an
```

possible reporting artefacts."

artifact P_outbreak_hypotheses : pattern(D_zone_time, M_transmission
representation form: "ranked origin and transmission hypotheses"
description: "Interpreted candidate origin zone-time areas, likely
propagation paths, leading mechanism explanation, alternatives,
residual anomalies, and explicit uncertainty."

transform T_assess_hypotheses :
intent: assess
manner: "evaluate evidential adequacy and actionability"
input: P_outbreak_hypotheses, V_hypotheses, K_epidemiology
output: K_hypothesis_quality
actor: human
description: "Judge whether source and mechanism inferences are st
epidemiologically plausible, and adequate for provisional action"

artifact K_hypothesis_quality : knowledge(P_outbreak_hypotheses)
representation form: "quality judgment"
description: "Assessment of confidence, unresolved alternatives,
sensitivity to assumptions, and whether refinement is required."

if K_hypothesis_quality indicates refinement needed:
then:

transform T_refine_outbreak_analysis :
intent: generate-knowledge
manner: "revise temporal scale, lag assumptions, thresholds,
contextual covariates, or mechanism comparison criteria"
input: K_hypothesis_quality, P_outbreak_hypotheses,
V_hypotheses, S_outbreak_analysis
output: S_outbreak_analysis_r
actor: human
description: "Refine the analysis configuration in response to
residual patterns, implausible assumptions, or unresolved
mechanism ambiguity."

artifact S_outbreak_analysis_r : specification
representation form: "revised outbreak-analysis settings"
description: "Updated aggregation, lag, spatial-neighbourhood,
mechanism-model, and evaluation settings."

```

        assign:
            S_outbreak_analysis := S_outbreak_analysis_r

    else:
        exit loop L_outbreak_hypothesis

end loop L_outbreak_hypothesis

# =====
# STEP 5: OPERATIONAL KNOWLEDGE SYNTHESIS
# =====

transform T_generate_response_guidance :
    intent: generate-knowledge
    manner: "formulate uncertainty-aware emergency recommendations"
    input: P_outbreak_hypotheses, K_hypothesis_quality, F_origin_scores,
           F_mechanism_fit, F_residuals, K_epidemiology
    output: K_response_guidance
    actor: human
    description: "Produce a ranked, conditional response plan: priority areas for field investigation and care deployment; mechanism-specific interventions; required confirmatory evidence; and surveillance activities"

artifact K_response_guidance : knowledge(P_outbreak_hypotheses,
    K_hypothesis_quality)
    representation form: "ranked findings, conditional recommendations, and uncertainty statement"
    description: "Operational guidance stating: likely origin area(s) and onset period; ranked transmission hypotheses and supporting evidence; unexplained anomalies; priority zones for clinical outreach and case investigation; conditional environmental sampling or airflow/water investigation; recommended additional data collection; and explicit caveat that microblog evidence alone cannot establish causality."

```