

Analytical logic

This workflow treats microblogs as a **syndromic proxy**, not as confirmed infection records. It first identifies messages likely to describe flu-like illness, then aggregates them into zone-by-day outbreak indicators normalized by population.

The analyst visually examines maps and time series to identify **candidate origin areas** based on the earliest *sustained excess* of illness-related activity—not simply the first isolated post. The workflow then compares alternative transmission explanations:

- **Person-to-person:** spread through adjacent or highly connected zones, population density, and transport corridors.
- **Airborne:** downwind and time-lagged propagation consistent with observed wind.
- **Water-associated:** clustering near water bodies or spatial patterns plausibly associated with water infrastructure.
- **Other/local source:** a localized point-source pattern near a landmark, facility, or event area.

The models provide comparative evidence rather than causal proof. The analyst iteratively inspects evidence, adjusts assumptions or lags, and may conclude that the available data cannot reliably distinguish transmission routes. The final output identifies candidate source locations, ranks transmission explanations with uncertainty, recommends resource deployment, and specifies additional data required for confirmation—especially confirmed case, death, contact, environmental, and water-quality data.

atwl

```
workflow vastopolis-flu-outbreak-investigation
  template: loop(characterise -> visualise -> abstract -> assess) ->
    define-unit -> define-unit -> characterise ->
    contextualise -> visualise -> abstract ->
    loop(build-model -> characterise -> visualise -> abstract ->
    generate-knowledge
  description: "Human-in-the-loop investigation of a suspected flu-like
    outbreak using geolocated microblogs as a syndromic surveillance pro
    The workflow identifies candidate origins and compares person-to-per
    airborne, water-associated, and other local-source explanations whil
    preserving uncertainty."

# =====
# INPUT DATA, CONTEXT, AND PRIOR KNOWLEDGE
```

```
# =====

artifact D_posts : entities
  origin: given
  internal structure: elementary
  embedment: {space, time}
  features:
    - id: timestamp
      value structure: atomic
      value type: temporal
      description: "Time at which the message was posted"
    - id: location
      value structure: atomic
      value type: spatial
      description: "GPS coordinate associated with the mobile device"
    - id: text
      value structure: atomic
      value type: text
      description: "Microblog message content"
  description: "Geolocated microblog messages posted during the suspected outbreak period"

artifact D_zones : entities
  origin: given
  internal structure: region
  embedment: space
  features:
    - id: zone_id
      value structure: atomic
      value type: reference
      description: "Identifier of one of the 13 metropolitan zones"
    - id: population
      value structure: atomic
      value type: numeric
      description: "Resident population of the zone"
    - id: geometry
      value structure: atomic
      value type: spatial
      description: "Zone boundary geometry"
  description: "Metropolitan zones used for population-normalised aggregation and emergency-resource planning"
```

```
artifact D_geography : entities
  origin: given
  internal structure: elementary
  embedment: space
  features:
    - id: geographic_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 geographic feature"
  description: "Reference map features, including 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 daily wind speed"
    - id: wind_direction
      value structure: atomic
      value type: spatial
      description: "Observed daily wind direction"
  description: "Daily weather observations during the investigation peri
```

```
artifact K_epidemiology : knowledge(D_posts, D_zones, D_weather)
  origin: given
  representation form: "domain rules and expected signatures"
  description: "Public-health knowledge about flu-like symptom language,
    plausible reporting biases, plausible transmission signatures, and t
    distinction between syndromic signals and confirmed infections"
```

```
artifact S_relevance : specification
  origin: given
  representation form: "method choice and thresholds"
  description: "Initial rules and model settings for identifying posts 1
```

to describe flu-like illness, including symptom terms, negation and language variants, confidence threshold, and exclusion criteria"

artifact S_hypothesis : specification

origin: given

representation form: "comparative modelling configuration"

description: "Initial temporal lags, spatial neighbourhood definitions candidate transmission mechanisms, uncertainty treatment, and criteria for comparing person-to-person, airborne, water-associated, and local-source explanations"

```
# =====  
# STEP 1: ITERATIVELY IDENTIFY ILLNESS-RELATED MICROBLOG POSTS  
# =====
```

loop L_relevance_refinement:

purpose: "Develop an interpretable and sufficiently reliable syndromic signal from unstructured microblog text"

until: "The analyst judges that illness-related messages are adequately distinguished from unrelated, historical, sarcastic, and non-local posts"

body:

transform T_score_posts :

intent: characterise

manner: "text encoding and illness relevance scoring"

input: D_posts, K_epidemiology, S_relevance

output: F_illness_score

actor: hybrid

description: "Extract symptom expressions, illness mentions, negation, severity language, and relevance confidence from post text"

artifact F_illness_score : feature(D_posts)

value structure: vector

value type: {numeric, categorical}

representation form: "illness relevance score and linguistic indicators"

description: "Per-post probability of describing a current local flu-like illness and associated symptom-language indicators"

transform T_visualise_posts :

intent: visualise

manner: "linked text and map inspection"

input: D_posts, F_illness_score

```
output: V_post_relevance
actor: machine
description: "Show sampled messages, relevance scores, locations,
posting times so the analyst can inspect false positives and omi

artifact V_post_relevance : visualisation(D_posts, F_illness_score)
layout: "linked message list, map, and time histogram"
form: "text records, coloured map points, and bars"
encoding: "colour indicates illness relevance score; map position
from GPS coordinates; time histogram shows posting volume"
description: "Interactive inspection view for validating the syndr
message-identification process"

transform T_interpret_language :
intent: abstract
manner: "perception-and-interpretation"
input: V_post_relevance, F_illness_score, K_epidemiology
output: P_language_patterns
actor: human
description: "Identify meaningful symptom vocabulary, ambiguous ph
reporting artefacts, and systematic false-positive message types

artifact P_language_patterns : pattern(D_posts, F_illness_score)
representation form: "annotated linguistic categories"
description: "Observed illness-language patterns and misleading me
categories relevant to flu-like syndromic surveillance"

transform T_assess_relevance :
intent: assess
manner: "judge classification adequacy"
input: V_post_relevance, P_language_patterns, F_illness_score
output: K_relevance_quality, S_relevance_r
actor: human
description: "Assess whether the relevance signal is sufficiently
credible for aggregate surveillance and, if needed, formulate
revised scoring rules or thresholds"

artifact K_relevance_quality : knowledge(F_illness_score, V_post_rel
representation form: "quality judgment"
description: "Analyst judgment about the adequacy, bias, and limit
of the illness-message relevance classification"
```

```

artifact S_relevance_r : specification
  representation form: "revised method choice and thresholds"
  description: "Updated linguistic rules, confidence threshold, or
    exclusion criteria for illness-message identification"

if K_relevance_quality indicates refinement needed:
  then:
    assign:
      S_relevance := S_relevance_r
    else:
      exit loop L_relevance_refinement

end loop L_relevance_refinement

# =====
# STEP 2: DEFINE OUTBREAK UNITS AND COMPUTE SPATIOTEMPORAL INDICATORS
# =====

transform T_select_illness_posts :
  intent: define-unit
  manner: "filter by illness relevance"
  input: D_posts, F_illness_score, S_relevance
  output: D_illness_posts
  actor: hybrid
  description: "Retain posts judged sufficiently likely to describe a
    current flu-like illness"

artifact D_illness_posts : entities
  internal structure: elementary
  embedment: {space, time}
  features:
    - id: illness_confidence
      value structure: atomic
      value type: numeric
      description: "Confidence that the post describes flu-like illness"
    - id: symptom_terms
      value structure: list
      value type: text
      description: "Extracted symptom and illness expressions"
  description: "Filtered geolocated posts forming the syndromic surveill
    signal for the outbreak investigation"

```

```
transform T_aggregate_zone_days :  
  intent: define-unit  
  manner: "space-time aggregation by zone and day"  
  input: D_illness_posts, D_zones  
  output: D_zone_days  
  actor: machine  
  description: "Aggregate illness-related posts into daily episodes for  
    city zone, retaining constituent messages for drill-down inspection"
```

```
artifact D_zone_days : entities  
  internal structure: episode  
  embedment: {space, time}  
  features:  
    - id: zone_id  
      value structure: atomic  
      value type: reference  
      description: "Zone represented by the zone-day episode"  
    - id: date  
      value structure: atomic  
      value type: temporal  
      description: "Calendar day represented by the episode"  
  description: "Zone-by-day syndromic surveillance episodes containing  
    illness-related microblog messages"
```

```
transform T_characterise_outbreak :  
  intent: characterise  
  manner: "population-normalised outbreak profiling"  
  input: D_zone_days, D_zones  
  output: F_outbreak_signal  
  actor: machine  
  description: "Compute illness-post counts, population-normalised rates  
    temporal changes, excess activity, and uncertainty caused by low vol
```

```
artifact F_outbreak_signal : feature(D_zone_days)  
  value structure: vector  
  value type: numeric  
  representation form: "daily syndromic incidence profile"  
  description: "Counts, rates per population, growth measures, anomaly s  
    and uncertainty indicators for each zone-day"
```

```
transform T_characterise_context :  
  intent: characterise
```

```
manner: "derive environmental and propagation relations"
input: D_zone_days, D_zones, D_geography, D_weather
output: F_exposure_context, F_propagation_relations
actor: machine
description: "Derive zone-day descriptors and lagged relations involving
population density, spatial adjacency, highways, hospitals, water
proximity, wind direction, and wind speed"
```

```
artifact F_exposure_context : feature(D_zone_days)
  value structure: vector
  value type: {numeric, categorical}
  description: "Population, geographic, weather, and infrastructure context
associated with each zone-day"
```

```
artifact F_propagation_relations : feature(D_zone_days)
  value structure: relational configuration
  representation form: "directed lagged zone-day relation graph"
  description: "Candidate propagation relations between zone-days based on
adjacency, transport connectivity, wind alignment, and water proximity"
```

```
# =====
# STEP 3: VISUALLY IDENTIFY CANDIDATE ORIGIN AREAS AND DIFFUSION PATTERNS
# =====
```

```
transform T_arrange_zone_days :
  intent: contextualise
  manner: "map-based"
  input: D_zone_days, D_zones
  output: A_zone_map
  actor: machine
  description: "Place each zone-day episode in its corresponding metropolitan
zone so time-filtered outbreak maps can be inspected"
```

```
artifact A_zone_map : arrangement(D_zone_days)
  context: D_zones
  principle: "zone identifier mapped to zone geometry"
  description: "Map arrangement of zone-day outbreak episodes"
```

```
transform T_visualise_situation :
  intent: visualise
  manner: "coordinated spatiotemporal outbreak display"
  input: A_zone_map, F_outbreak_signal, F_exposure_context,
```

```

        D_geography, D_weather
output: V_outbreak_situation
actor: machine
description: "Display animated and time-linked maps, zone time series,
        wind vectors, population-normalised rates, highways, hospitals, and
        water bodies"

artifact V_outbreak_situation : visualisation(A_zone_map, F_outbreak_sig
        F_exposure_context, D_geography, D_weather)
layout: "linked metropolitan map, zone time series, and temporal overv
form: "choropleth regions, proportional symbols, line graphs, and wind
encoding: "zone colour and symbol size show population-normalised illn
        activity; time controls select dates; overlays show geographic featu
        and observed wind direction"
description: "Interactive situational-awareness display for locating e
        sustained activity and visually inspecting possible spread pathways"

transform T_abstract_observed_patterns :
    intent: abstract
    manner: "perception-and-interpretation"
    input: V_outbreak_situation, F_outbreak_signal, F_propagation_relation
    output: P_observed_diffusion
    actor: human
    description: "Identify candidate source zones, timing of sustained exc
        wave-like spread, spatial clusters, anomalies, and patterns inconsis
        with a single simple outbreak explanation"

artifact P_observed_diffusion : pattern(D_zone_days, F_outbreak_signal)
    representation form: "candidate foci and spatiotemporal pattern descri
    description: "Analyst-identified candidate origin areas and observed
        diffusion structures, explicitly distinguished from confirmed source

# =====
# STEP 4: ITERATIVELY COMPARE TRANSMISSION EXPLANATIONS
# =====

loop L_hypothesis_comparison:
    purpose: "Compare plausible transmission mechanisms and determine whet
        the available evidence supports a defensible operational conclusion"
    until: "A best-supported explanation, or a justified conclusion that t
        data cannot distinguish mechanisms, has been reached with documented
        uncertainty"

```

body:

```
transform T_build_mechanism_models :  
  intent: build-model  
  manner: "fit competing spatiotemporal transmission models"  
  input: F_outbreak_signal, F_exposure_context, F_propagation_relati  
         P_observed_diffusion, K_epidemiology, S_hypothesis  
  output: M_mechanism_models  
  actor: hybrid  
  description: "Fit and compare models representing person-to-person  
               diffusion, wind-associated airborne spread, water-associated  
               clustering, and localized or other point-source mechanisms"
```

```
artifact M_mechanism_models : model(F_outbreak_signal,  
  F_exposure_context, F_propagation_relations)  
  model type: "comparative spatiotemporal explanatory models"  
  representation form: "set of competing lagged diffusion and source  
  description: "Alternative formal explanations of observed zone-day  
               illness activity; model comparison measures evidence rather than  
               proving a causal transmission route"
```

```
transform T_characterise_model_evidence :  
  intent: characterise  
  manner: "compute model fit and diagnostic evidence"  
  input: D_zone_days, F_outbreak_signal, F_exposure_context,  
         F_propagation_relations, M_mechanism_models  
  output: F_hypothesis_evidence  
  actor: machine  
  description: "Calculate comparative fit, residuals, uncertainty,  
               predicted propagation, and local evidence for each transmission  
               hypothesis"
```

```
artifact F_hypothesis_evidence : feature(D_zone_days)  
  value structure: vector  
  value type: numeric  
  representation form: "hypothesis fit, residual, and uncertainty pr  
  description: "Zone-day evidence indicating how well competing  
               explanations account for observed syndromic activity"
```

```
transform T_visualise_hypothesis_evidence :  
  intent: visualise  
  manner: "comparative model diagnostics display"
```

```

input: A_zone_map, F_outbreak_signal, F_hypothesis_evidence,
      M_mechanism_models, D_geography, D_weather
output: V_hypothesis_evidence
actor: machine
description: "Compare observed and predicted spread, residual maps
temporal lags, downwind alignment, water proximity, and transport
connectivity in coordinated views"

artifact V_hypothesis_evidence : visualisation(A_zone_map,
      F_outbreak_signal, F_hypothesis_evidence, M_mechanism_models,
      D_geography, D_weather)
layout: "linked observed-versus-predicted maps, timelines, and
hypothesis comparison panels"
form: "choropleths, flow arrows, residual marks, line graphs, and
comparative score plots"
encoding: "map panels show observed and model-predicted zone activity;
arrows show candidate propagation relations; colour and residual
marks show local model agreement or disagreement"
description: "Interactive evidence display used to inspect whether
apparent transmission signatures are coherent, robust, and plausible"

transform T_interpret_mechanism_evidence :
  intent: abstract
  manner: "perception-and-interpretation"
  input: V_hypothesis_evidence, F_hypothesis_evidence,
        K_epidemiology
  output: P_mechanism_evidence
  actor: human
  description: "Interpret model evidence in public-health context;
identify support, contradictions, confounding, and uncertainty for
each candidate transmission mechanism"

artifact P_mechanism_evidence : pattern(M_mechanism_models,
      F_hypothesis_evidence)
representation form: "ranked mechanism evidence statements"
description: "Interpreted regularities supporting or weakening
person-to-person, airborne, water-associated, or local-source
explanations"

transform T_assess_hypotheses :
  intent: assess
  manner: "evaluate explanatory adequacy and operational usefulness"

```

```

input: V_hypothesis_evidence, P_mechanism_evidence,
      M_mechanism_models, K_epidemiology
output: K_hypothesis_assessment, S_hypothesis_r
actor: human
description: "Judge whether model assumptions, evidence, uncertainty
and visual explanations are adequate for a response decision;
formulate revised lags, constraints, or comparison settings if n

artifact K_hypothesis_assessment : knowledge(M_mechanism_models,
      P_mechanism_evidence)
representation form: "quality judgment and uncertainty assessment"
description: "Assessment of whether the evidence adequately distin
mechanisms, identifies a plausible candidate source, and support
emergency action"

artifact S_hypothesis_r : specification
representation form: "revised comparative modelling configuration"
description: "Updated temporal lags, spatial relations, constraint
or robustness checks required for another hypothesis-comparison

if K_hypothesis_assessment indicates further comparison needed:
then:
    assign:
        S_hypothesis := S_hypothesis_r
else:
    exit loop L_hypothesis_comparison

end loop L_hypothesis_comparison

# =====
# STEP 5: PRODUCE OPERATIONAL FINDINGS AND DATA-REQUEST RECOMMENDATIONS
# =====

transform T_generate_response_knowledge :
intent: generate-knowledge
manner: "formulate-statements and recommendations"
input: P_observed_diffusion, P_mechanism_evidence,
      K_hypothesis_assessment, K_relevance_quality,
      M_mechanism_models, D_zones, D_geography
output: K_outbreak_assessment, K_response_recommendations,
      K_confirmation_data_needs
actor: human

```

description: "Formulate a public-health assessment that states candidate origin zones, ranked transmission explanations, confidence and caveats, recommended deployment actions, and the additional evidence needed to validate the conclusions"

artifact K_outbreak_assessment : knowledge(P_observed_diffusion,
P_mechanism_evidence, M_mechanism_models)

representation form: "qualified situational assessment"

description: "Candidate outbreak origin area or areas, timing of emergence, ranked transmission explanations, supporting evidence, contradictory evidence, and explicit uncertainty"

artifact K_response_recommendations : knowledge(K_outbreak_assessment,
D_zones, D_geography)

representation form: "prioritised emergency-response recommendations"

description: "Recommended deployment of testing, clinical capacity, communications, environmental investigation, and surveillance resources conditional on the best-supported mechanism and affected zones"

artifact K_confirmation_data_needs : knowledge(K_outbreak_assessment)

representation form: "data-collection recommendations"

description: "Required confirmation data, such as confirmed case and death line lists, laboratory results, contact information, mobility data, water-quality samples, water-network information, and environmental sampling results"