University of St. Gallen

School of Computer Science

to obtain the title of

Master of Science in Computer Science

Master Thesis on

At the Edge of Empathy

Mental-Health Awareness Through Home-Router Network Traffic Insights

---

Submitted by:

Stephan Nef

18-652-180

Approved on the application by:

Prof. Dr. Bruno Rodrigues

and

Prof. Dr. Simon Mayer

Date of Submission:

October 31, 2025

Acknowledgements

I am truly grateful to my master's thesis advisor, Prof. Dr. Bruno Rodrigues, for his constant support, motivation and inspiring guidance throughout this work.

I would also like to thank Prof. Dr. Simon Mayer for co-supervising my master's thesis.

My sincere thanks also go to my family and close friends for their unwavering encouragement, support and patience with me during this interesting but also intense time. I am equally grateful to my manager at work and colleagues for their flexibility, which enabled me to complete this work alongside my professional commitments.

St. Gallen, October 31, 2025 Stephan Nef

Abstract

At the Edge of Empathy explores how home network routers, often regarded as silent intermediaries of digital life, can become privacy preserving instruments for mental health awareness. Positioned at the edge of the network, the router offers a unique vantage point that is close enough to capture behavioral rhythms while distant enough to protect the content of personal communication. This balance between technical precision and human sensitivity frames the central idea of the thesis, which is empathy through design rather than intrusion. Depression often develops gradually, but the behavioral changes that precede a diagnosable episode are rarely observed early enough for effective intervention. Device centered sensing can be intrusive and fragmented across devices, which leaves a gap for transparent household level observation. To address this gap, we design, implement, and evaluate a router centric system that transforms home network metadata into interpretable behavioral indicators aligned with the definition of major depressive disorder in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5). We formalize Behavior Observation Metrics for mental health domains such as sleep disturbance, loss of interest, and concentration difficulty, with local processing and full transparency. Network traces are divided into five minute windows, enriched and mapped to per criterion likelihoods through a Fuzzy Additive Symptom Likelihood model. Parameters such as weights, membership functions, thresholds, and gate windows are explicit and user configurable. An interactive dashboard supports daily status tracking and criterion level exploration. An implemented application integrates packet capture, feature extraction, fuzzy aggregation, and an auditable DSM-5 style decision gate. The evaluation shows that we map multiple header-based metrics to Behavior Observation Metrics (BOM), and that the aggregated BOM indicate to the respective DSM-5 criterion for depressive disorder. The results are explainable and reproducible over multiple days, without requiring payload access. Flow level signals such as night to day ratios, domain diversity, and the share of passive versus active traffic covary with target DSM-5 criteria. The resulting end to end, privacy preserving architecture connects digital phenotyping with transparent network analytics and establishes a replicable and auditable foundation for ethically responsible early awareness of depression.

Table of Contents

List of Figures

List of Tables

List of Algorithms

List of Listings

List of Equations

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
APNs	Apple Push Notification service
ASN	Autonomous System Number
BOCPD	Bayesian Online Change Point Detection
BRAS	Broadband Remote Access Server
CDN	Content Delivery Network
CSV	Comma-Separated Values
CUSUM	Cumulative Sum (change detection)
CV	Coefficient of Variation
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DFI	Deep Flow Inspection
DHCP	Dynamic Host Configuration Protocol
DPI	Deep Packet Inspection
DNS	Domain Name System
DOI	Digital Object Identifier
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EM	Expectation–Maximization
EPDS	Edinburgh Postnatal Depression Scale
ETL	Extract, Transform, Load
eTLD+1	Effective Top-Level Domain plus one

EWMA Exponentially Weighted Moving Average

FCM Firebase Cloud Messaging

GAM Generalized Additive Model

GMM Gaussian Mixture Model

GPS Global Positioning System

HDRS Hamilton Depression Rating Scale

HOB Human-Observable Biomarker

HTTP Hypertext Transfer Protocol

HTTPS Hypertext Transfer Protocol Secure

IAB Interactive Advertising Bureau

ICD-10 International Classification of Diseases, Tenth Revision

IKI Inter-Keystroke Interval

IKS Inter-Keystroke Interval (seconds)

IP Internet Protocol

ISP Internet Service Provider

JSON JavaScript Object Notation

LSTM Long Short-Term Memory

MAC Media Access Control

MDD Major Depressive Disorder

MDE Major Depressive Episode

ML Machine Learning

MVP Minimum Viable Product

NTP Network Time Protocol

ODN	Optical Distribution Network
OLT	Optical Line Terminal
ONU	Optical Network Unit
PCAP	Packet Capture
PHQ-9	Patient Health Questionnaire 9
Pi-hole	Pi-hole DNS Sinkhole
PON	Passive Optical Network
PSL	Public Suffix List
QUIC	Quick UDP Internet Connections
SCAM	Shape-Constrained Additive Model
SERP	Search Engine Results Page
SI	Suicidal Ideation
SLD	Second-Level Domain
SNI	Server Name Indication
STL	Seasonal–Trend decomposition using Loess
TCP	Transmission Control Protocol
TLS	Transport Layer Security
UDP	User Datagram Protocol
UI	User Interface
VPN	Virtual Private Network
ZScore	Standard Score (z-score)

Chapter 1

Introduction

Depression is a leading contributor to the global burden of disease and often emerges during adolescence, where it forecasts poorer educational achievement, strained peer relationships, and an elevated risk of unemployment in adulthood [33, 12]. Although effective psychotherapeutic and pharmacological treatments exist, their success relies on timely initiation. Meta-analytic evidence shows that every additional month of untreated illness in youth lowers the chance of full remission [33]. Observational cohort studies in high-income settings still report median delays of several months between symptom onset and first evidence-based treatment [37]. Early intervention therefore depends on recognizing behavioral change well before a formal depressive episode satisfies diagnostic criteria. Recent advances in ubiquitous and unobtrusive sensing such as smartphone usage logs, keyboard dynamics, ambient light or motion detectors make it feasible to capture minute-to-minute variations in sleep, activity, social connectedness, and circadian regularity in real-world settings [45, 57].

1.1 Motivation

Leveraging continuous, passively collected metadata can move assessment beyond sporadic self-report toward an objective, high-resolution behavioral baseline against which clinically meaningful deviations can be detected in near real time. At the same time, actionable insight does not require fine-grained content or invasive collection; coarse, device-level and environmental features are often sufficient to indicate pattern shifts relevant to depressive behavior. Installing apps across devices that continuously capture and export data raises privacy and ethical concerns, as well as deployment burden. Therefore, this thesis explores a router-centric approach that confines collection to network metadata at a single, consented observation point, links signals to individuals within a shared household, and processes data locally whenever feasible.

Research Questions

1. Can a router centric pipeline reliably attribute device level and environmental signals to a single individual in a multi person household?
2. Do these signals, when aggregated, correlate with indicators of depressive behavior?
3. Can these indicators enable parents, caregivers, and clinicians to observe early indicators of depression while preserving privacy and meeting ethical safeguards?

Consider a multi-person household in which all devices connect through a single home router. Over several weeks, one smartphone exhibits later first-packet-of-day times, a larger share of nocturnal bursts and in future could also be leveraged with data like co-presence events recorded by doorway presence sensors and ambient light. This thesis links such network traffic data to a single individual in a shared environment and summarizes them over *Behavior Observation Metrics (BOM)* toward DSM-5 Criterion Indicators to indicate depressive behavior mapped to a single individual. The scenario grounds the problem throughout the thesis while keeping message content private and limiting data collection.

1.2 Clinical and Technical Background

Current diagnostic manuals, namely the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) and the *International Classification of Diseases, Tenth Revision* (ICD-10), define symptom criteria for a Major Depressive Episode [39, 53]. Screening tools such as the Patient Health Questionnaire–9 (PHQ-9) translate these criteria into brief self-report scales, but they depend on accurate retrospective recall [26]. Validation studies show that recall bias and postponed help-seeking lead to a substantial share of episodes being missed at first contact [47].

In response to that, research has turned to passive digital phenotyping. Passive digital phenotyping can be described as the process of identifying and characterizing observable features of an organism, within this thesis a Human, through passive digital sensing. Smartphones and wearables record mobility traces, screen-on events, call logs, or sleep estimates that correlate with depressive symptom severity [45, 58]. However, these studies often assume that a participant continuously carries one dedicated device and runs a research app throughout the study. In everyday family settings these assumptions break

down: devices are shared, batteries are depleted, or the user disables an app precisely when motivation declines [60]. Missing data then match with the period that is most informative from a clinical perspective.

Previous studies revealed that WiFi metadata can serve as a non-intrusive proxy for behavioral patterns relevant to mental health [59, 31]. Social-media analyses reached similar conclusions by examining linguistic content [11, 14, 8]. Recent advances in domain labeling with the Public Suffix List (PSL) and the IAB Content Taxonomy suggest that clinically relevant information can be extracted from headers alone, without invading textual privacy [34, 23].

1.3 Research Gap

Digital phenotyping has advanced rapidly, yet four identified open research gaps still limit its clinical impact.

1. **G1: Short behavioral excursions are missed** Most phone-based studies compress their data into daily or weekly averages [45, 58]. Early depressive signals, however, often appear as a single night of fragmented sleep or one day of social withdrawal. When such brief events are averaged with surrounding days, they disappear. Clinicians then learn about the change only after compensatory routines, e.g. weekend catch-up sleep, have blurred the pattern [19].

2. **G2: Household traffic is a convoluted mixture** Routers see packets from siblings, parents, guests, and televisions. Without reliable device or person level separation the subtle changes that matter for one adolescent are diluted by the stable habits of others, which lowers sensitivity and raises false alarms [59, 31].

3. **G3: Header level dynamics remain under used** Packet headers contain information about burst timing, protocol variety, and destination categories. These cues mirror sleep wake rhythm, task switching, and shifting interests, yet prior work usually stops at simple byte counts [59]. Recent taxonomies such as the Public Suffix List and the IAB Content Taxonomy show that richer, privacy preserving labels are available but still largely unexplored [34, 23].

4. **G4: Opaque models slow clinical adoption** Machine learning systems based on deep embeddings or large ensembles often achieve high accuracy [60], but do not

explain sufficiently transparent why a warning was triggered. Clinicians often need a clear link between a model characteristic and a DSM-5 criterion before acting on an indication towards a DSM-5 MDD criterion. [53, 39]. Without that link, promising algorithms might remain academic prototypes.

Closing these gaps would support a shift from reactive care, which treats full episodes months after onset, to proactive support that helps families and clinicians intervene during the earliest detectable drift towards disorder [33, 37].

1.4 Methodological Overview

Early detection requires a transparent path from raw packets to clinical insight. Therefore, this thesis adopts a workflow (Figure 1.1) in which each layer performs a distinct task and hands a clean output to the next, while every DSM-5 MDD criterion indicator rests on an explainable case.

1. **Digital Environment** Continuous passive capture at the household router delivers a minute by minute record of device activity without asking users to install apps or fill in diaries, an approach that has already proven feasible and privacy aware in home WiFi studies [59, 31].

2. **Feature Extraction** The packet stream is compressed into numerical descriptors such as first online time, burst length, traffic diversity and repetition rate and mapped to Behavior Observation Metric (BOM) and aggregated into DSM-5 MDD Criteria Indicator. Previous work shows that these low level features retain behavioral signals that can predict future mood change while cutting the data volume by several orders of magnitude [45, 58].

3. **Behavior Observation Metric (BOM)** Related features are grouped into behavior centred markers, for example reduced social interaction, flattened activity rhythm or repetitive browsing loops. Bundling features in this way reduces noise and yields concepts that also non specialists can understand. It is a strategy found to be effective in keyboard- and phone based phenotyping [64, 46].

4. **Clinical Mapping to DSM-5 MDD criteria** Each BOM contributes to the indicator towards a DSM-5 MDD criterion it best represents, such as e.g. social withdrawal or sleep disturbance. Linking data driven markers with formal criteria makes

indications interpretable for parents or clinicians and supports evidence based intervention [39, 53].

1.5 End-to-End Workflow

Figure 1.1 shows the full user journey in one view. We begin with upload and profiling on the left, proceed to exploratory plots that surface baselines and weekly regularities, then move to per-metric calibration with interpretable memberships and limits. The lower panels show how calibrated metrics aggregate into criterion-level likelihoods and, finally, how the DSM-5 gate summarizes the rolling window for an episode-level overview. We return to each step in depth: the metric design and fusion logic in Chapter 4, the data flow and dashboards implemented in Chapter 5, and the empirical evaluation with criterion traces and gate behavior in Chapter 6.

Figure 1.1: End-to-end workflow. The application guides the user from dataset selection and user profile setup to exploratory baselines, per-metric parameter tuning, criterion-level indicators as day-level likelihoods and a DSM-5 gate with a rolling window and core-symptom rule.

This teaser anchors the storyline for the remaining chapters. We design the interpretable metrics and aggregation once, we implement them in a reproducible router-centric system, and we evaluate what the plots reveal about daily behavior and criterion likelihoods. The same visual language appears again in the Methodological Framework for formal defini-

tions, in the Implementation chapter for the concrete user interface (UI) and data pipeline, and in the Evaluation chapter for results and limitations under realistic constraints.

1.6 Scientific and Practical Significance

This thesis contributes a fresh lens on digital phenotyping by showing that network data can be translated into clinically meaningful indicators of depression. Scientifically, this demonstrates an unexplored, platform-agnostic data source and offers a concise mapping from packet headers to DSM-5 MDD criteria, that future mental-health studies can replicate or extend. Practically, the approach can be deployed with commodity routers and privacy-sensitive processing, allowing households to obtain early indicators without wearables, apps, or other more intrusive solutions. In short, the work advances both the methodological toolkit for mental-health research and the feasibility of real-world, low-cost screening.

1.7 Scope

In this thesis we develop and examine a router-centric approach that converts home-network data into transparent, daily indicators aligned with the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5). The privacy-first pipeline confines observation to packet-header and metadata at a single, consented gateway, (ii) formalizes an approach with Behavior Observation Metric (BOM) that groups related features into interpretable components, (iii) designs and implements a Minimum Viable Product (MVP), specifically a web application that computes daily, criterion-level likelihoods using fuzzy memberships and an explicit DSM-style gate, and (iv) evaluates correctness, interpretability, and limitations on realistic, longitudinal network traces.

The scope centers on adolescents in multi-person households, where attribution and privacy are critical. The vantage point is the home router and payloads should not be inspected. Analyzes focus on header-level timing and coarse domain labels to derive per-criterion likelihoods, daily status views, and reproducible configuration artifacts. The work aims to help parents, caregivers and clinicians to observe early behavioral change through auditable, non-diagnostic indicators and to provide a methodological foundation other studies can replicate or extend.

1.8 Contributions

This thesis addresses the open gaps, outlined in Section 1.3 (G1–G4) by designing, implementing, and evaluating the above mentioned router-centric, privacy-preserving and interpretable pipeline for behavioral observation aligned with DSM-5. The contributions are:

1. **Router-centric, privacy-preserving data capture and attribution** A household-gateway vantage point confines observation to non-content metadata while enabling person-level attribution in multi-user settings. The approach builds on the foundations for networking and privacy in Chapter 2, Sections 2.2 and 2.3. The implemented architecture in Chapter 5, Sections 5.2, 5.3, 5.4, and 5.4.1 operationalizes this capture in a reproducible way. This directly tackles privacy-preserving collection and household attribution (G2, G3).

2. **Behavior Observation Metric (BOM) with criterion-aligned design** Features extracted are grouped into interpretable *Behavior Observation Metrics (BOM)* that correspond to clinically meaningful patterns. Criterion-aligned metrics are specified in Chapter 4ff., ensuring each metric has a clear label, risk direction, and justification. This addresses interpretable feature mapping to clinically grounded criteria (G3, G4).

3. **Transparent likelihoods and auditable decision logic** Daily criterion likelihoods are computed with explicit membership functions and additive fusion in Section 4.5, with shapes in Section 4.5.1 and alternatives in Section 4.5.2. Multi-metric fusion is defined in Section 4.5.3. A DSM-style gate then converts likelihood histories into criterion and episode decisions using simple, verifiable rules, see Section 4.5.4, Figure 4.20. The full path from raw packets to decisions remains inspectable. This addresses transparent inference and auditable decision-making (G1, G4).

4. **End-to-end prototype with interactive dashboards** A functional system integrates data ingestion, feature extraction, and user-journey visualization for non-technical stakeholders. Chapter 5 details the UI semantics and status formatting (Section 5.5.1), profile handling (Section 5.7), and the user journey (Section 5.8). This addresses feasibility and stakeholder-oriented interpretability at router scope (G2, G4).

5. **Evaluation on realistic network traces with emphasis on interpretability** Chapter 6 assesses the robustness of the metric pipeline and the clarity of the resulting indicators on longitudinal traces, with examples that illustrate how short behavioral excursions and longer trends surface. The evaluation focuses on transparency and operational feasibility consistent with the design goals. This addresses empirical validation under realistic constraints (G1, G3).

Together, these contributions position a privacy-first, router-centric pipeline that yields interpretable, DSM-aligned indicators without payload inspection. The approach is situated within the state of the art in Chapter 3 and carried through the design in Chapter 4 and the implementation in Chapter 5, culminating in the empirical assessment in Chapter 6.

1.9 Thesis Outline

Table 1.1 provides a compact storyline of the thesis. It highlights the router-centric, privacy-preserving, and interpretable approach, and points to the exact chapters and sections where each element is developed and assessed.

Table 1.1: Thesis Outline Overview

Chapter	Topics	Key References
Introduction (Chapter 1): Overview of motivation, scope, and contributions.	Motivation and problem framing, research gap, scope and contributions.	[33, 12, 37]
Background (Chapter 2): Clinical and networking foundations.	Networking foundations, what metadata remains observable and privacy, ethics and compliance.	[39, 45, 57, 42]
Related Work (Chapter 3): Gaps and positioning in the literature.	Survey of passive sensing and network-level inference, limitations of device-centric sensing and positioning of a router-centric, transparent approach.	[59, 60, 49]
Methodological Framework (Chapter 4): Interpretable metrics and decision logic.	Criterion-aligned metrics and BOM formulation with fuzzy memberships and additive fusion, transparent multi-metric fusion and a DSM-style gate.	[5, 49]
Implementation (Chapter 5): System, data pipeline, and dashboard.	Implemented architecture, data sources, PCAP ingestion and five-minute partitioning, DNS and domain enrichment, formatting and status cues, user profiles and dashboard navigation.	[61, 6, 34, 51]
Evaluation & Discussion (Chapter 6): Criterion likelihoods and DSM-5 gate.	Bottom-up metric calibration and criterion-level fusion with contribution traces and the DSM-5 gate	[58, 11, 45]
Conclusion & Future Work (Chapter 7): Synthesis, lessons learned and future next steps.	Synthesis, lessions learned and limits of the router-centric approach and a roadmap for labeled field studies and explainable auto-tuning.	[30, 17, 1, 62]

Chapter 2

Background

Depression is difficult to diagnose in its early stages. Formal diagnostic criteria in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) and the *International Classification of Diseases, Tenth Revision* (ICD-10) are typically applied only after an individual seeks professional help [47, 39]. This approach misses opportunities for early detection because it relies on self-reported symptoms, retrospective recall, and delayed clinical contact [53].

2.1 Clinical Frameworks for Depression

DSM-5 defines a Major Depressive Episode (MDE) as the presence of at least five out of nine core symptoms, for a minimum of two consecutive weeks [53]. The core symptoms include depressed mood, anhedonia, sleep disturbance, or fatigue. At least one of the five symptoms must be depressed mood or loss of interest/pleasure. These symptoms are typically measured through structured clinical interviews, standardized rating scales (e.g., Patient Health Questionnaire, PHQ-9), and patient self-report questionnaires that systematically assess the frequency, duration, and severity of each symptom. The manual thus combines a quantitative threshold ($\geq$5 symptoms) with two qualitative clauses: duration (two weeks) and functional impact, evaluated through these standardized clinical assessment tools.

2.1.1 DSM-5 Criteria for Major Depressive Episode

- Criterion 1: Depressed mood (subjective report or observed by others).
- Criterion 2: Markedly diminished interest or pleasure (anhedonia) in all or almost all activities.
- Criterion 3: Significant weight loss or gain (e.g., >5% body weight in a month) or decrease/increase in appetite; in children, failure to make expected weight gain.

- Criterion 4: Insomnia or hypersomnia.
- Criterion 5: Psychomotor agitation or retardation (observable by others).
- Criterion 6: Fatigue or loss of energy.
- Criterion 7: Feelings of worthlessness or excessive/inappropriate (possibly delusional) guilt.
- Criterion 8: Diminished ability to think or concentrate, or indecisiveness.
- Criterion 9: Recurrent thoughts of death, recurrent suicidal ideation (with or without a plan), or a suicide attempt.

Additional requirements are (a) the symptoms cause clinically significant distress or impairment in *social, occupational, or other important areas of functioning*. (b) The episode is *not attributable* to the physiological effects of a substance or another medical condition. [55]

ICD-10 lists the same nine symptoms but distinguishes three core symptoms (depressed mood, loss of interest, and reduced energy) from seven additional symptoms. Severity is graded by simple symptom counts. Mild episodes show at least four symptoms and include at least two core symptoms. Moderate episodes require six symptoms and also include at least two core symptoms. Severe episodes require eight or more symptoms, encompass all three core criteria, and may include psychotic features [39]. These symptoms are typically assessed through structured clinical interviews and standardized screening tools such as the Patient Health Questionnaire (PHQ-9), Hamilton Depression Rating Scale (HDRS), or clinician-administered diagnostic assessments. Thus, the ICD scheme provides a more nuanced clinical vocabulary while maintaining the same minimum two-week duration criterion as DSM-5 [26].

Although both standards bring diagnostic consistency, they also introduce practical limitations for early depression detection:

1. **Episodic Snapshots** Diagnosis is triggered only when patients seek help or when a periodic screening is performed. In primary care, up to 50 % of depressive episodes are missed at first contact [39].
2. **Retrospective and/or Subjective Self-Report** Scales validated against DSM-5 or ICD-10, such as the Edinburgh Postnatal Depression Scale [47], rely on the patients

recall of the past two weeks, which can be biased or poorly calibrated in younger populations.

3. **Static Duration Threshold** The mandatory two-week window may overlook short yet clinically relevant mood swings. Adolescents in particular can experience rapid changes linked to school pressure, social media exposure, or sleep disruption [2].

4. **Implications for Sensor Research** To complement, rather than replace, DSM-5, a passive sensing platform must (i) detect behavioral deviations at a finer temporal scale, (ii) operate unobtrusively in a multi-user environment, and (iii) aggregate features into the *Behavior Observation Metric (BOM)* that maps onto DSM-5 criteria.

5. **Subthreshold Presentations in Childhood and Adolescence** Many young people display only two to four depressive symptoms for days or weeks too few, or for too short a time, to satisfy the five-symptom, two-week rule in DSM-5. These subthreshold episodes can already affect school performance and peer relationships and increase the risk of a later Major Depressive episode. In routine primary care, such brief or mild presentations often go unnoticed. Up to half of all depressive cases are missed at first contact [39]. Therefore, a continuous, sensor-based approach offers a practical path to indicate early behavioral change.

2.2 Foundations of Home-Network Traffic Capture

When a device in a household connects to the internet, each transmission generates technical information that is not hidden by encryption. Even if the message contents are encrypted using protocols such as Transport Layer Security (TLS 1.3) or QUIC, the router still observes basic metadata: who sent the packet, where it is going, how large it is and when it was sent. These metadata provide a continuous picture of activity without exposing private content.

The home gateway occupies a unique position. Before all devices are merged into a single public Internet Protocol (IP) address through Network Address Translation (NAT), it can still distinguish between individual devices inside the household. This ability is important for attributing activity to a specific user or device, while shielding those identities from the outside world. Not all traffic is visible in this way. If devices use a Virtual Private Network (VPN), mobile connections, or certain privacy extensions, their activity may

bypass the home router. In such cases, these missing observations must be carefully treated in the design and evaluation of models based on network data.

2.2.1 Content Inference from Encrypted Traffic

Although payloads are encrypted, some information remains accessible. The Server Name Indication (SNI), exchanged during the TLS handshake, and the plain-text Domain Name System (DNS) request that precedes a DNS-over-HTTPS (DoH) session can still reveal the fully qualified domain name. By mapping domains to the Mozilla Public Suffix List [34] and then to the Interactive Advertising Bureau (IAB) Tech Lab Content Taxonomy [23], we can derive coarse, privacy-preserving content labels. Examples include `/Games/Online` or `/Social_Networks`. For example, a spike of requests to `*.gaming.net` at 03:00 can be a machine-readable hint of disturbed sleep. Disturbed sleep is one DSM-5 symptom domain for Major Depressive Disorder (MDD).

2.3 Privacy, Ethics and Legal Compliance

This thesis follows a privacy-by-design approach that limits observation to network and environmental metadata and avoids message content. We align with the *Federal Act on Data Protection* (FADP) in force since 1 September 2023 for work conducted in Switzerland, adopt high-level privacy-by-design requirements from *ISO/IEC 31700-1:2023*, and consider the obligations that would apply under the *European Union Artificial Intelligence Act* (AI Act) if the system were deployed in clinical or caregiver settings.[18, 24, 42] Under the FADP, packet-header and sensor metadata that can be linked to an identifiable individual constitute personal data [18]. Processing therefore requires a clear purpose, data minimization, and appropriate safeguards [18]. For households with minors, participation requires parental consent and adolescent assent. We predefine a limited set of features, restrict collection to a single, consented observation point (the router), prefer local processing, pseudonymize device identifiers at capture, and enforce short retention windows. A *Data Protection Impact Assessment* (DPIA) could be prepared prior to any pilot with human participants.

2.3.1 Privacy by Design and Security Controls

Following ISO/IEC 31700-1:2023, it is foreseen that the methodological framework implements privacy by design across the lifecycle. Therefore, a clear purpose specification, local preprocessing and aggregation, separation of identifiers from feature data,

role-based access control, encryption in transit and at rest and documented retention and deletion procedures should be complied with the FADP and ISO/IEC 31700-1:2023 standard. [24, 18]. All Behavior Observation Metric (BOM) computations should be designed to be interpretable, and every component remains traceable to its underlying metadata features to support transparency and user rights. Also important to mention here is that the methodological framework should not inspect payload content, should not attempt clinical diagnosis and it should also not replace professional assessment.
In sum, a router-centric, privacy-first architecture fits this problem well. A single, consented observation point, the home router, enables strict data minimization and local preprocessing. Per-device attribution supports individual-level BOM computation in multi-person homes. Metadata-only features preserve content privacy while remaining interpretable and mappable to DSM-5 criteria, consistent with ISO/IEC 31700-1:2023, the Swiss FADP, and the EU AI Act [24, 18, 42].

The next section separates these safeguards from the question of how we treat BOM as a measurement instrument, with explicit expectations for validity and reliability.

2.4 Behavior Observation Metric as a Measurement Instrument

The Behavior Observation Metric (BOM) in this thesis is a composite indicator system that maps observable network and environmental features to criterion-aligned evidence for DSM-5 MDD symptom domains. The use of BOM is not a diagnostic tool. It aggregates feature-level evidence into interpretable, criterion-level likelihoods to support earlier observation and discussion.

2.4.1 Validity Expectations

We aim for (i) *content validity*, by aligning each metric with a DSM-5 MDD criterion through an explicit rationale; (ii) *face validity*, by ensuring that each metric and its 'risk direction' are understandable to clinicians and caregivers; and (iii) *construct validity*, by checking whether groups of metrics that are intended to reflect the same symptom domain covary as expected. These expectations are operationalized through transparent membership functions (MF), per-metric weights, and a DSM-style gate that are all documented and auditable.

2.4.2 Reliability Expectations

We expect *temporal stability* when behavior is stable and *responsiveness* when behavior changes. Practically, we probe short-term stability with repeated daily windows and we check sensitivity to change around known perturbations (e.g., sleep disruptions). Thresholds and weights are calibrated conservatively to favor interpretability and stability over black-box accuracy.

2.4.3 Transparency by Design

All parameters required to compute BOM-feature definitions, MFs, weights, aggregation rules, and the criterion gate-are specified in the methodological framework (Chapter 4) and evaluated in Chapter 6. This allows external review and downstream clinical validation without exposing message content. Chapter 4 details the operationalization: how constructs map to features, how MFs and weights are set, and how we handle multi-person households. Chapter 6 reports reliability probes and construct-validity checks using longitudinal windows and criterion-aligned comparisons.

Next, we review related work, mostly done on mobile and network-based sensing (Chapter 3) to clarify what is known and where limitations persist. Consistent with the research gaps stated in Chapter 1, Section 1.3, Chapter 4 introduces a router-centric methodological framework designed for privacy-preserving attribution and interpretable indicators aligned with the DSM-5 standard.

Chapter 3

Related Work

Building on the theoretical foundations introduced in Chapter 2, this section situates the present and relevant studies within the wider body of research on passive, sensor-based depression monitoring. To make the landscape tractable the existing work is mapped into four complementary strands. Clinical symptom frameworks, device-centric digital sensing, environmental sensing via network-traffic metadata, and multimodal fusion across heterogeneous sensors. This taxonomy not only mirrors the data sources leveraged by our system but also exposes methodological blind spots that motivate the thesis contributions. The Table A.1 summarizes representative studies in each strand, flags the research gaps (G1–4) distilled from the literature, and foreshadows how the forthcoming chapters address them.

Building on Chapter 2, we situate this thesis within passive, sensor-based depression indication. To make the landscape tractable we group prior work into four strands:

1. **Clinical Symptom Frameworks (DSM-5, ICD-10)**: This review analyzed established diagnostic frameworks, emphasizing their inherent limitations such as the requirement for symptoms to persist for a minimum duration of two weeks, thereby delaying timely interventions. More details in Section 2.1, critical evaluations of clinical practices and diagnostic thresholds were informed by literature from Pedersen et al. [39], Smith-Nielsen et al. [47], and Svenaeus [53]. In this work, we focus on the DSM-5 standard, however, the scope of our approach is not limited to it.

2. **Passive Digital Sensing**: Research on digital phenotyping via smartphones and wearables was systematically reviewed, see also Section 3.1, emphasizing their potential for unobtrusive and continuous detection of depressive symptoms. Significant contributions reviewed included studies by Wang et al. [58], who explored

mobile and wearable sensing among college students, and Saeb et al. [45], who demonstrated the correlation between mobile phone sensor data and depressive symptom severity. However, most studies either average over long windows (research gap G1) or assume a one-person–one-device scenario (research gap G2).

3. **Environmental Sensing with an Emphasis on Network Traffic Metadata**: This strand examined literature on using network-level metadata for environmental sensing, see also Section 3.2, highlighting how passive network metadata can effectively reveal behavioral indicators linked to depressive states without compromising individual privacy. Seminal works reviewed include Ware et al.'s analysis of WiFi infrastructure metadata for depression screening [59], and Mammen et al.'s passive WiFi sensing approach to detect sleep disturbances associated with depression [31].

4. **Fusing Environmental and Device-Centric Views**: Finally, the review considered literature on multimodal sensor fusion, emphasizing the advantages of integrating environmental sensors with personal digital devices. Specifically, it highlighted how network traffic metadata enables effective assignment and alignment of sensor events across devices and individuals, thereby improving detection accuracy and interpretability in shared household environments [59, 31].

3.1 Passive Digital Sensing in Mental-Health Research

Digital phenotyping refers to automated collection of behavioral signals from everyday devices such as smartphones, tablets, or wearables. These devices generate a continuous stream of metadata that can reveal clinically relevant patterns without active user input. However, most studies either average over long windows (G1) or assume a one-person–one-device scenario (G2).

3.1.1 Multi-Sensor Hubs

Stachl et al. showed that behavioral metadata from consumer smartphones can predict broad and facet-level Big-Five traits across more than 25 million logging events [49]. Two lessons map to this thesis. First, high-level psychological constructs can, in principle, be inferred from low-level traces, which supports the face validity of header-only features. Second, the authors released only engineered features, not raw logs, due to privacy concerns, mirroring our decision to keep processing on the household gateway.

Stachl et al. demonstrated in a large-scale mobile-sensing study that purely behavioral metadata harvested from consumer smartphones can predict both broad and facet-level Big-Five personality traits with considerable accuracy across over 25 million logging events [49]. Two lessons from their work map directly on the present work. First, the result confirms that high-level psychological constructs can, in principle, be inferred from low-level digital traces, underscoring the clinical potential of our own header-only network features.
Second, the authors released only extracted features and never the raw logs because of privacy concerns (see also Chapter 2, Section 2.3 about Privacy, Ethics and Legal Compliance), mirroring our decision to keep all packet processing on the household gateway. By extending Stachl et al. "sensor-rich smartphone" paradigm to a router-centric, multi-resident environment, this thesis seeks to achieve similar indication power while further reducing participant burden and privacy risk. Another Pioneering work such as *StudentLife* demonstrated that a single mobile application can infer sleep, mobility, and social rhythms in college students and link these rhythms to self-reported mood [57]. Later studies showed that features derived from accelerometer variance, GPS radius of gyration, call logs, and screen-on events explain up to 40% of the variance in depressive symptom severity [45, 58]. All of these signals originate from devices that must exchange packets with the surrounding networking infrastructure to stay synchronized. In that way they are able to fetch push notifications or simply back up sensor data to the cloud. Consequently, the underlying networking traffic already contains coarse proxies for many of the same behaviors (for example, the timing and size of push notification bursts can signal screen engagement [27]).

These findings motivate router-side metadata as proxies for the same constructs while addressing G1 and G2. However, high-performing pipelines often rely on opaque embeddings or ensembles, which weakens clinical interpretability G4 [58, 45].

3.1.2 Wearables and App-Centric Studies

Sleep trackers, smartwatches, and fitness bands add heart-rate and inertial measurements with higher fidelity than phones can typically provide. The *SleepTracker* application, for instance, detected anxiety and depression related sleep disturbances in an sample [2]. Although such wearables can operate offline, most commercial devices periodically upload summaries to vendor servers. The timestamps and byte counts of these uploads

appear as distinctive flows in the home router's logs, making them observable without installing software on the wearable itself.

Sleep trackers, smartwatches, and fitness bands provide heart-rate and inertial measurements with higher fidelity than phones. The *SleepTracker* app detected anxiety- and depression-related sleep disturbances in an outpatient sample [2]. Router-side observation captures these uploads as metadata events, supporting Behavior Observation Metric (BOM) related to sleep without additional burden.

3.1.3 Social Media Footprints

Another branch of research mines textual content from online social networks to detect linguistic markers of depression or suicidal ideation [11, 14, 8]. Social media interaction is therefore visible at the gateway as a series of short Transmission Control Protocol (TCP) or Quick UDP Internet Connections (QUIC) flows. Each post, like or scroll event creates a new flow whose header still exposes the destination domain. Transport Layer Security (TLS) encrypts the payload, yet it leaves the Server Name Indication field in clear text. That field is enough to resolve the service name with the *Public Suffix List* and to label the flow with a coarse content category defined by the *IAB Content Taxonomy* [34, 23].

Because no packet payload is inspected, the approach respects privacy constraints. Only the timing, direction and byte count of flows are stored. These metadata already capture meaningful behavior. Frequent mid-night requests to social-media domains point to sleep disruption, which is one of the core DSM-5 MDD symptoms. Rapid, repetitive bursts of traffic from the same handset suggest rumination or anxious checking behavior. Prior studies have linked such patterns to rising depressive symptom scores [11, 14].
Header-level packet data can contribute to BOM components aligned with sleep and activity patterns while maintaining privacy (addresses research gap G3).

3.1.4 Networking Metadata

Despite reliance on connectivity, systematic use of network traffic for mental-health sensing remains rare. Ware et al. captured association logs from enterprise access points and found that students with higher depressive scores exhibited fewer handovers and lower overall data volume [59]. Beyond such aggregate statistics, richer descriptors-flow regularity, protocol mix, and coarse domain taxonomy-remain underused (G3). This

thesis positions networking traffic as a unifying lens that passively reflects behavior across all devices used by an adolescent without installing apps or decrypting content.
The present thesis therefore positions networking traffic as a unifying lens: it passively reflects activity from *all* devices carried by a child or adolescent, and it does so without requiring apps to be installed or personal content to be decrypted.

3.1.5 Privacy-relevant User Profiling

Parallel to our router-centric, protocol-agnostic vantage point, recent work shows that even a single smart-home protocol can enable privacy-relevant user profiling. Müller et al. demonstrate a passive, inference-based tool that recognizes devices and events in ZigBee (Philips Hue) networks with accuracy of $\approx 94\%$ without decrypting payloads. While this focus is narrower (link-layer ZigBee only) than our all-traffic router vantage point, it reinforces both the privacy stakes and the feasibility of header/metadata-based inference in households [35]. In contrast, our approach aim for metadata-only capture at the home router (no payload inspection), local preprocessing with short retention and pseudonymized per-device attribution. From this protocol-agnostic vantage, features are mapped to DSM-5 MDD criteria through BOM with documented MFs, weights, and a DSM-style gate to support transparent, privacy-preserving user profiling in multi-person households.

3.1.6 Limitations of Device-Centric Sensing

Most phone-based studies presuppose that participants keep their handset nearby and continue to run a dedicated research application [45, 57, 58]. Both assumptions often fail when motivation declines or when depressive symptoms intensify; missing data then coincide with the period that is clinically most informative [60]. In households with several residents the problem grows. Packets that leave the gateway merge traffic from phones, laptops and game consoles belonging to different people. Without proper separation the extracted features form an aggregate that cannot be mapped back to a specific child or adolescent, which raises the risk of misclassifications [59, 31].
In this thesis, networking metadata will play a dual role, they provide behavioral evidence that can be aligned with the DSM-5 and ICD-10 symptom domains, and they serve as the "glue" that links different types of sensors while allowing coarse device identification without deep packet inspection. This joint view allows us to separate co-located users, raise temporal coverage when a child switches between devices, and extract Behavior

Observation Metric (BOM) that are robust enough for early detection in everyday family life. Within this thesis, networking metadata play a dual role. They provide behavioral evidence that can be mapped to DSM-5 MDD criteria, and they serve as the 'glue' that links sensors while allowing for the identification of the coarse device without the deep inspection of the packets. This joint view helps separate co-located users and raises temporal coverage when a child switches devices. The features are aggregated into the *Behavior Observation Metric (BOM)*, designed to be interpretable and robust for early observation in everyday family life.

In summary, passive digital sensing now covers a wide spectrum. Early studies used a single research app. Recent work mines millions of social-media posts. However, the networking layer that links every device in the household remains underexplored. Packets already carry timestamps, flow sizes and domain names. These data appear without any extra software on the phone or watch. They can therefore fill blind spots when a user disables an app or switches devices.

The next subsection shifts the focus to environmental sensing and how networking metadata can merge with signals such as room presence. The fusion allows us to answer three key questions: *who* generated the traffic, *where* at home it happened, and *when* during the day it occurred.

Taken together, device-centric sensing demonstrates predictive value but remains tied to single-user assumptions and opaque mappings. We therefore summarise these trade-offs against network-centric strands in a compact comparison that concludes this chapter (Section 3.2.3).

3.2 Environmental Sensing and Sensor-Fusion

Clinical assessment of depression already takes the patients environment into account. Physicians ask about sleep–wake rhythm, day-light exposure, physical activity and social withdrawal because each topic maps to one or more DSM-5 MDD symptom domains [53, 39]. However, the process is based on self-report and brief in-clinic observation. Continuous and objective measurement is rarely feasible in routine care. Low-cost infrared and ultrasonic sensors can log room occupancy, while photodiodes capture the daily light dose and such data reveal changes in sleep timing, social isolation or psychomotor slowing-three core criteria of a Major Depressive Episode. Mobile studies that record ambient audio have linked reduced speech duration to rising depression scores,

although privacy concerns limit large-scale deployment [45]. The sleep-tracker literature shows a similar trend. Deviations in total sleep time or sleep onset latency correlate with anxiety and depression severity [2].

3.2.1 Networking Metadata as an Environmental Sensor

Networking metadata turn the household router into a silent observer that never looks at payload yet still mirrors daily routines. Each packet header shows a time-stamp, a byte count and the source Media Access Control address. Successive association and disassociation events outline the path a phone or tablet takes from bedroom to kitchen and back. Ware et al. analysed such logs on a university campus and saw that students with higher depressive scores connected to fewer access points during the day and moved less after dusk [59]. The pattern matches two DSM-5 MDD criteria: psychomotor retardation and insomnia. Mammen et al. later counted gaps between packets at home gateways and inferred sleep duration without any devices on the body [31]. Their estimates captured nightly awakenings that correspond to the DSM-5 MDD criterion for 'disturbed sleep'. A follow-up study by the same group combined flow volume with self-reports and showed that decreases in night traffic preceded an increase in patient health questionnaire scores by about a week [60]. These findings support router-side attribution and BOM components for sleep and mobility while maintaining content privacy (addresses research gaps G1–3).

3.2.2 Interpretability and Clinical Adoption

Clinical decisions value measures that are transparent and traceable to symptom domains. Many digital-phenotyping models optimize accuracy using embeddings or stacked learners, but they provide limited rationale for individual alerts, which slows adoption in clinical workflows (G4) [53]. In contrast, this thesis aggregates metadata features into BOM's that are named, inspected, and explicitly mapped to DSM-5 constructs. This design choice prioritizes interpretability over black-box gains when they conflict with clinical usability.

Across the four strands, prior work establishes the feasibility of passive sensing but leaves recurring gaps: short behavioral excursions are averaged away (G1), attribution in multi-person homes is unresolved (G2), rich header-level metadata remain underused (G3), and opaque models slow clinical adoption (G4). Table A.1 shown in the Appendix

lists the most relevant studies, marks the gaps (G*n*), and anticipates our contributions (C*n*). Building on these take-aways, the next Chapter 4 specifies the methodological framework: a router-centric attribution pipeline and interpretable Behavior Observation Metric (BOM) components operationalized over short, fixed windows, along with the study design and analyses that will test attribution, validity, detection against baselines, and interpretability.

Infrastructure logs and gateway traces provide privacy-preserving observables, yet most studies stop at aggregate signals or single-behavior exemplars. The following synthesis (Section 3.2.3) contrasts these strands and positions our router-centric contribution before Chapter 4.

3.2.3 Discussion and Positioning

Table 3.1 contrasts the works surveyed in this chapter against six criteria central to this thesis: (i) *router centric* (home gateway), (ii) *metadata-only* (no payload), (iii) *person attribution* in multi-user households, (iv) *transparent/ interpretable* logic, (v) *DSM-style gate* for criterion decisions, and (vi) *privacy by design*. Each row corresponds to a cited work and carries a superscript that points to detailed notes in Appendix A (Table A.1).

Overall, no prior work covers all six dimensions. Device-centric strands offer temporal coverage but assume single-user devices and often rely on opaque predictors. Content-based social media methods compromise privacy. Existing network-centric studies operate at an aggregate level or focus on a single behavior. The next chapter introduces a router-centric, metadata-only pipeline with explicit attribution, transparent metric logic, and a DSM-style gate (Chapter 4), and Chapter 5 realizes it end to end.

Table 3.1: Comparison of related work vs. this thesis across centric, data, attribution, interpretability, DSM-style gate, and privacy. All superscripts refer to Appendix A with all notes about related work in Table A.1.

Reference	Router Centric	Metadata-only	Person attribution	Transparent / interpretable	DSM-style gate	Privacy by design
StudentLife: assessing mental health, aca...d behavioral trends of college students using smartphones (2014) [57]	No	No	(Yes)[1]	(Yes)[1]	No	No
Mobile phone sensor correlates of depressive s...tom severity in daily-life behavior: an exploratory study (2015) [45]	No	No	(Yes)[2]	No	No	No
Tracking depression dynamics in college students using mobile phone and wearable sensing (2018) [58]	No	(Yes)[3]	(Yes)[3]	(Yes)[3]	No	(Yes)[3]
Scalable passive sleep monitoring using mobile phones: opportunities and obstacles (2017) [46]	No	No	(Yes)[4]	No	No	No
"Silence your phones" Smartphone notifications increase inattention and hyperactivity symptoms (2016) [27]	No	(Yes)[5]	(Yes)[5]	(Yes)[5]	No	(Yes)[5]
Evaluating the effectiveness of the Sl...tecting anxiety-and depression-related sleep disturbances (2024) [2]	No	No	(Yes)[6]	(Yes)[6]	No	No
Early detection of depression: social network analysis and random forest techniques (2019) [11]	No	No	(Yes)[7]	No	No	No
Discovering shifts to suicidal ideation from mental health content in social media (2016) [14]	No	No	(Yes)[8]	(Yes)[8]	No	No
Leveraging LLM-generated data for detecting depression symptoms on social media (2024) [8]	No	No	(Yes)[9]	No	No	No
Large-scale automatic depression screening using meta-data from wifi infrastructure (2018) [59]	(Yes)[11]	Yes[11]	(Yes)	(Yes)[11]	No	(Yes)[11]
Wisleep: Inferring sleep duration at scale using passive wifi sensing (2021) [31]	Yes[12]	Yes[12]	(Yes)[12]	(Yes)[12]	No	Yes[12]
Predicting depressive symptoms using smartphone data (2020) [60]	Yes[13]	Yes[13]	No	(Yes)[13]	No	Yes[13]
Predicting personality from patterns of behavior collected with smartphones (2020) [49]	No	(Yes)[10]	(Yes)[10]	(Yes)[10]	No	(Yes)[10]

all superscripts point to the per-work notes in Appendix Table A.1.

Chapter 4

System Design

This chapter specifies our router-centric design that turns network data into daily indicators aligned with the DSM-5 criteria introduced in Section 2.1. We want to keep computation and storage on the household gateway. The idea is that the router observes all devices in one place and provides a stable point for feature extraction and decision logic which keeps processing efficient, secure, central and lightweight. We build on established blocks from networking and measurement rather than creating new ones. Our contribution is to combine them into a transparent, explainable and configurable pipeline that fits the constraints of a home network and the privacy principles set out in Section 2.3. The foundations for packet capture and domain enrichment are summarized in Section 2.2 and Section 2.2.1, including the Public Suffix List and the IAB content taxonomy that we can reuse for service grouping [34, 23]. Prior work shows the value of passive, device-level signals for behavior assessment in daily life, which motivates a router vantage point for household-level sensing, also see in Table 3.1 [57, 45, 59, 60, 31].

4.1 Architecture Overview

We first state the edge boundary of our system. Figure 4.1 shows the perimeter of privacy. Devices generate traffic that reaches the home router. Feature extraction, explainable scoring and secure storage run inside this layer. The output is a set of daily DSM-5 indicators. User telemetry is not sent outside the home. This boundary drives all later decisions on features, scoring and configuration.

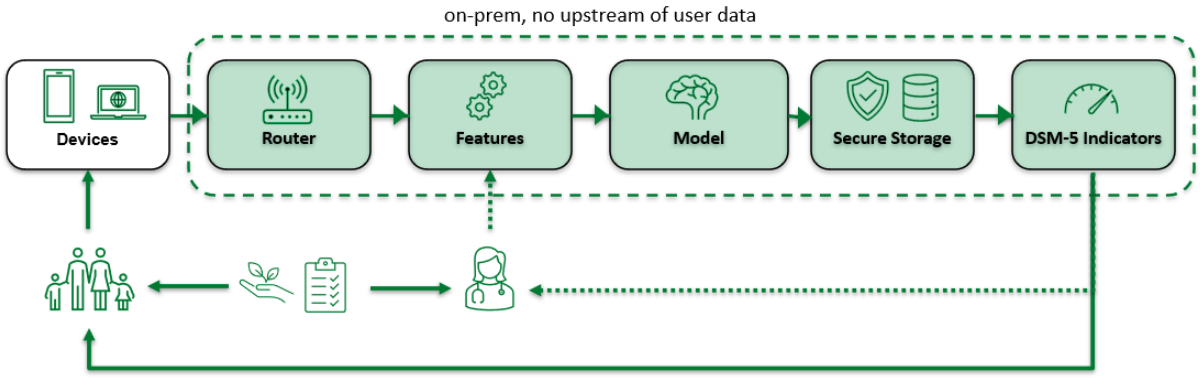


Figure 4.1: Edge workflow and privacy boundary. Processing remains on-premises inside the dashed box. On-device feature extraction, explainable scoring, and secure storage produce daily DSM-5 indicators with no upstream transfer of user data.

At block level, the pipeline follows a simple sequence that we reuse throughout the thesis. Raw traces are ingested, normalized and enriched with domain context where possible. Robust daily and weekly features are computed. Each feature contributes to a Behavior Observation Metric (BOM) for one DSM-5 criterion. BOM are combined into criterion-level likelihoods with explicit auditable membership functions. A DSM-Gate maps these likelihoods to day-level indicators. Profiles capture all thresholds and weights so that results remain reproducible and explainable between devices and households.

4.2 Design Objectives & Paradigms

The system was developed to achieve four objectives that serve as guidelines for feature extraction, grouping, aggregation and decision rules. In line with our general approach of optimizing proven ideas rather than reinventing the wheel, the objectives and corresponding paradigms draw on established network and data protection practices and build on previous findings from edge-based behavior recognition. see Section 2.2 ff. and Section 2.3, [59, 60, 31, 57, 45, 49, 34, 23].

4.2.1 Conceptual Design Objectives

O1 Non-intrusive and privacy-preserving meaning to use data at a single, consented observation point and prefer local processing, minimize data retention. This follows the privacy-by-design principles referenced in Section 2.3 and aligns with current guidance and regulation [24, 18, 42].

O2 Clinically aligned at the DSM-5 symptom level meaning to map observable signals to DSM-5 MDD criteria by computing interpretable metrics and BOM components for each criterion. We rely on the clinical framing introduced in Section 2.1 and reuse routine- and rhythm-oriented signals documented in prior work on passive, everyday sensing [57, 45, 59, 60, 31, 49].

O3 Transparent and configurable meaning to expose memberships, weights, thresholds and gate parameters by providing per-criterion explanations and configuration export/import for reproducible results. Parameterization and explanations are specified in Section 4.5 ff., and implemented as profile settings in Chapter 5.

O4 Robust to missing or sparse data meaning to tolerate gaps, flag insufficient history and prefer personalized baselines. This choice reflects the household edge setting discussed in Section 4.1 and is consistent with the stability advantages of daily/weekly aggregates reported in [59, 60, 31].

4.2.2 Conceptual Design Paradigms

P1 A transparent end-to-end workflow from raw PCAP windows to day-level, DSM-5–aligned symptom indicators with explicit, adjustable steps. The design is instantiated in a multi-page dashboard that computes daily features and render per-criterion explanations and reference ranges. This end-to-end chain mirrors the edge workflow in Figure 4.1 and the more detailed System Design later in Figure 4.2.

P2 A simple, expressive aggregation layer that maps heterogeneous metrics, that are each assigned to a BOM, to a daily likelihood for each DSM-5 MDD criterion (by using triangular membership functions with explicit weights, followed by an interpretable DSM-style gate parameterized by window M, required count N, and threshold θ, explained in Section 4.5 ff.) This keeps scoring explainable and lightweight for router-grade hardware.

P3 The use of BOM groups as interpretable sets of related network-traffic features. BOM groups prevent single-feature dominance and make model logic easier to discuss with clinicians. They act as labels and structure for weighting and future variants such as clustering, not as standalone metrics. The feature families and service groupings reuse established traffic categories and domain reductions from Section 2.2.1 [34, 23].

P4 A configurable and auditable setup in which the full model can be exported and imported, with an optional fully explainable auto-tune step (e.g. a possible fuzzy k-means clustering) that proposes membership parameters and weights from historical baseline data in a reproducible manner. The configuration and audit trail are part of the profile mechanism in Chapter 5, and the auto-tune operates strictly within the explainable rules defined in Section 4.5.

4.2.3 Privacy-first Edge Design Choice

The household router is the natural place to operate. It already concentrates device traffic, which provides complete coverage without installing software on personal devices. This vantage point captures routine behavior while keeping payloads opaque and personal content private, as already discussed in Section 2.3.1. Processing and storage remain on the router. We minimize retention and keep identifiers separated from features. Means, identifiers such as MAC addresses or host names are kept in a small, access-controlled table, while the feature store contains only pseudonymous device tokens with time-stamped aggregates. A temporary audited join is created only when rendering a user profile and removing the mapping makes the historical features non-attributable to a person, which supports the privacy controls. The design follows the privacy by design and security controls described in Section 2.3 and is in line with current regulatory expectations [24, 18, 42]. To mitigate common risks, external observation is limited by transport encryption. Local access to the device could be protected by standard router hardening and encrypting stored profiles. Any change to the settings could be noticed, because all parameters are listed clearly, when saved with a version/session number, and linked to a profile. This approach works on standard home routers and keeps the system lightweight.

4.2.4 Edge-Centred Operating Envelope

This section explains the practical choices that let the system run smoothly on a home router while protecting privacy. We size the pipeline for modest hardware so that it works reliably alongside normal internet use. We keep it portable by using common router interfaces. And we build signals from privacy-preserving metadata, so the results remain useful in a clinical context without inspecting content. Finally, we describe what runs on the home gateway and why keeping processing inside the home helps with both trust and efficiency.

4.2.4.1 Design Envelope and Interoperability

We assume a home gateway with modest CPU, memory, and storage capacity. The design therefore focuses on header-based metadata, one-pass feature derivations, and compact daily or weekly aggregates to keep processing on the router light and continuous without compromising its primary function. Scoring uses simple, closed-form membership functions and an interpretable gate, which aligns with the explainability and efficiency goals set out in Section 4.5 ff.. This resource-aware approach matches the edge boundary introduced in Section 4.1 and the privacy principles in Section 2.3, and is consistent with previous edge-centric sensing pipelines [59, 60, 31].

Interoperability is addressed at the design level by relying on standard observation points and broadly supported capture interfaces so that different router models and firmware can participate without custom instrumentation. Domain and service grouping reuse established resources such as the Public Suffix List and content taxonomies, allowing the enrichment layer to be refreshed as data resources evolve while the core logic remains unchanged [34, 23]. Thresholds, windows and weights are kept in profile configurations, which makes the system portable across households while preserving a single, auditable logic as detailed later in Chapter 5.

4.2.4.2 Inference Accuracy and Data Scope

We use metadata and traffic rhythm rather than payload content. This choice favors privacy and generalisability. The features describe activity bursts, diurnal patterns, and service categories. BOM translate these features into criterion-level signals using interpretable rules, then the DSM-Gate in Section 4.5.4 turns them into daily indicators. The scope fits early signs where broad changes in routine and engagement matter more than message content [59, 60, 49]. Later chapters evaluate how far this approach reaches and where additional sensors could complement the router view without changing the overall design.

4.2.4.3 Computation and Storage on the Household Gateway

We keep computation and storage on the home gateway because it is the single place that already observes all devices in a home, and it allows us to apply one privacy policy for everyone. The gateway runs the full pipeline from capture to indicators within the dashed boundary shown in Figure 4.1. Capture and parsing operate on header information only as described in 2.2. Domain names are reduced locally to registrable eTLD+1 labels using the Public Suffix List and optional IAB categories from 2.2.1. Features and criterion likelihoods are computed in short, fixed windows so that memory stays bounded

and intermediate data can be deleted after use, which aligns with the privacy controls summarized in Section 2.3. The same design appears in the implementation where we partition traces into five minute files and derive a single daily record before scoring, see Section 4.2.4, 4.1 and the find the corresponding implementation in Chapter 5 . This keeps processing efficient on router-grade hardware and keeps data inside the household.

In practice, two deployment forms work well. When the router can host user workloads the analytics service runs directly on the device and writes encrypted local stores with short retention, as outlined in Section 2.3. When the router is closed, a small companion box inside the same home network receives a mirror of network traffic and performs the same local steps. In both cases identifiers are stored separately from features and configuration is versioned so that every indicator can be traced back to its inputs, see Section 4.5.4.There is prior evidence that this edge approach is feasible and useful. WiFi infrastructure logs and home-gateway traces have been used to infer sleep timing and duration as well as mobility and routine changes without looking at payloads [59, 31, 60]. These works motivate the router vantage and support our decision to keep processing local while mapping signals to transparent metrics and daily indicators defined in Chapter 4.

4.3 System Overview

In Figure 4.2 we summarize the end-to-end workflow and reference the components as they appear in the diagram. Traffic from all connected devices on the home network first reaches *Home Router* and is mirrored at `Gateway`, where `Network Capture (PCAP / tcpdump)` records packet information. An `ETL Parser & Filtering` stage extracts timestamps, addresses, ports, protocol bytes, and, when available, DNS or SNI hostnames (not limited to it). Under the privacy-first regime indicated by the dashed boundary, records are organized in `Partitioned Storage` as fixed segments of n-minutes that allow local processing, data minimization, and short retention before any analysis proceeds.

In parallel, the system maintains an inventory of devices via *Device & Session Discovery* and resolves household ambiguity through `Identity Resolution (IP to Profile allocation)`, storing assignments in `User Profile Registry`. These identity services feed the analytical path at two points. First, they inform `Feature Extraction` so that the measurements per-segment reflect the intended individual, and second, condition `Feature Store`, where the records per-segment are retained. A lightweight

Feature Cache accelerates iterative analysis and a Parameter Registry maintains analyst-controlled settings for memberships, weights, and gate parameters.

From the feature store, the measures aligned with the criterion are calculated in Feature Engineering and grouped into interpretable components within BOM.
The DSM-5 MDD Criterion Modules (C1-C9) transform these components into daily probables based on a possible criterion $L_k \in [0,1]$. The DSM Gate then applies a clinically motivated rule on a rolling window M: a criterion is marked present when $L_k \geq \theta$ on at least N days. However, the design that is not restricted to this mechanism means that any other transparent and clinically interpretable model that fulfills the requirements could be integrated at this stage.
An overall Depression Episode Likely indicator is reported when at least five criteria are present and at least one core criterion is present as defined by the DSM-5 standard. All filter settings, configuration, calculations, and results are presented in the Interactive Indicator Dashboard, which provides explanations per-criterion, BOM traceability, and auditable configuration so that the results can be reproduced at any time.

The figure also marks cross-cutting properties. The privacy-first lane covers capture, storage and ETL so that network traffic data are processed, identifiers are separated from features, and retention is bounded. The personalisation lane indicates that analyst-set parameters and per-user weights enable profile-aware thresholds while keeping outputs interpretable. Together, identity resolution, fixed-window storage, engineered header signals, BOM components, and the explicit gate link the home vantage point to daily clinical indications in a way that is transparent, configurable, and aligned with the DSM-5 standard.

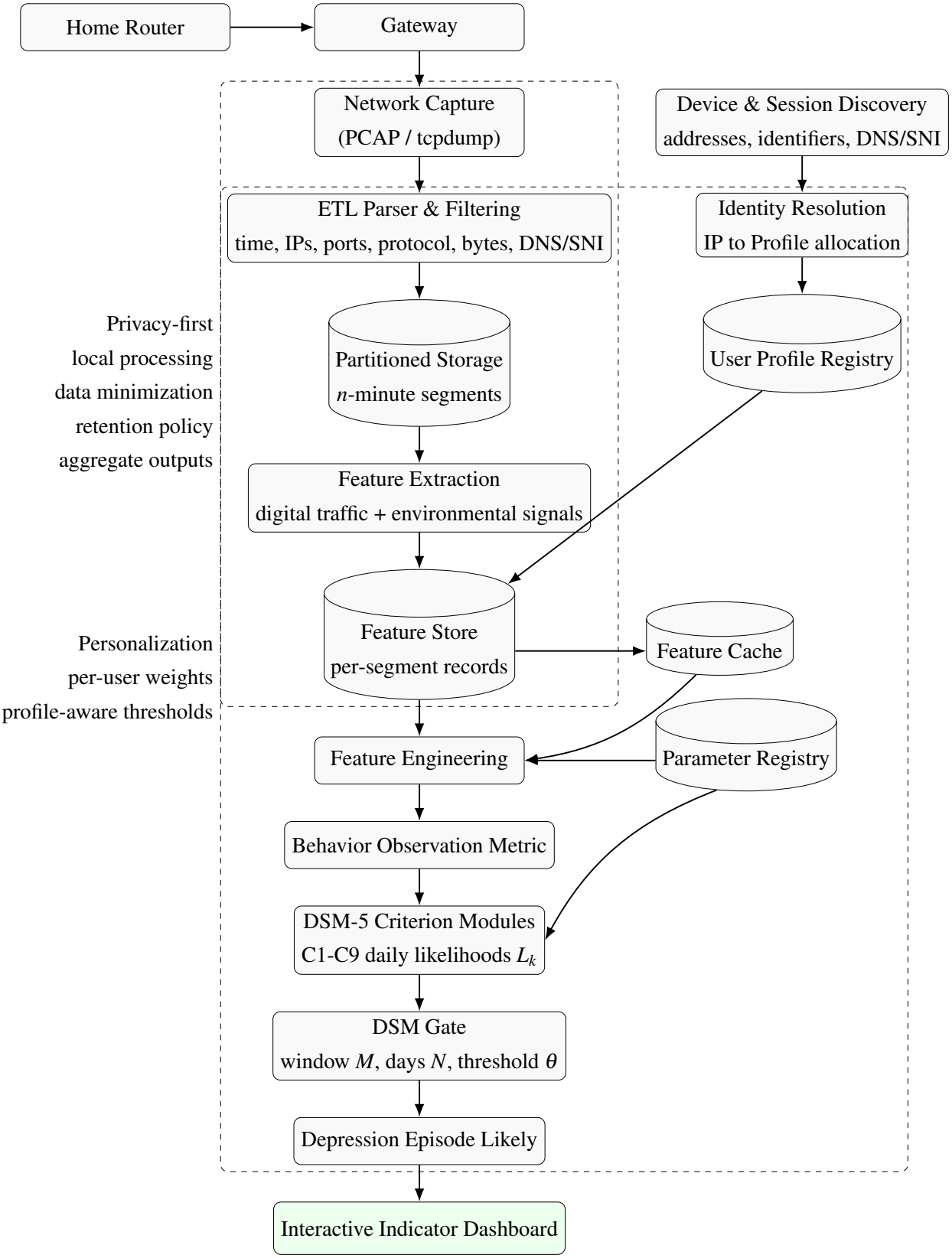


Figure 4.2: System Design shows the router-centric capture flow through parsing, storage, feature extraction, BOM towards DSM-5 MDD indicators and a gate yield an episode-likely indicator.

4.3.1 Layered Mapping to DSM-5 Indicators

In this subsection, we introduce the layered framework that links everyday household network activity to interpretable indicators aligned with the DSM-5 MDD criteria for a depressive episode. The flow visualized in Figure 4.3 illustrates how the raw packet data collected at the home router gradually becomes structured information that can support clinical reasoning. The layers are arranged from left to right, beginning with the data source and ending with clinical interpretation. Each layer aggregates, filters, or translates information from the previous layer while keeping privacy and interpretability in focus. The numerical computations that connect these layers are described in more detail later on.

4.3.1.1 Layer 1: Data capture at the home router

Household routers provide a single, privacy-preserving observation point for all connected devices. Traffic is captured passively at the router interface as packet-header data, without requiring software on individual devices. This approach keeps personal content off the router, yet preserves timing and structural features that reflect digital behavior. At this layer, the observation unit is the household, which forms the basis for later aggregation into daily or hourly PCAP windows.

4.3.1.2 Layer 2: Feature extraction and temporal partitioning

The captured packets are processed into sessions and summarized over fixed time windows. In this thesis, daily windows are used, but the same framework can operate on hourly or multi-day windows. Each window produces a structured set of network features such as traffic volume, protocol composition, burstiness, idle gaps, and upload-download balance that describe household activity patterns. This temporal partitioning makes it possible to trace gradual behavioral changes over time rather than relying on absolute values.

4.3.1.3 Layer 3: Mapping from metrics to Behavior Observation Metrics (BOM)

Individual traffic features are grouped into higher-level Behavior Observation Metrics (BOM), which represent interpretable behavioral aspects observable through network activity. For instance, irregular nighttime traffic may inform a BOM for sleep–wake rhythm, while interaction patterns across devices may inform social engagement. The relationship between metrics and BOM is many-to-many: one metric can contribute to several BOM, and one BOM can be informed by multiple metrics.

4.3.1.4 Layer 4: Alignment with DSM-5 MDD criteria

Each BOM is aligned with one or more DSM-5 MDD criteria for a Major Depressive Episode (Criteria 1–9). This alignment does not diagnose. It estimates how strongly a behavioral pattern corresponds to a given symptom domain such as loss of interest, disturbed sleep, or psychomotor change. The mapping allows several BOM to contribute jointly to one criterion, reflecting the multifactorial nature of mental health symptoms. The outcome is a set of likelihood-style indicators that summarize the presence or absence of specific criteria within each observation window.

4.3.1.5 Layer 5: Clinical assessment and feedback

The resulting indicators are intended to support, not replace, professional interpretation. They can be reviewed by clinicians, caregivers, or parents alongside established self-report instruments such as the PHQ-9. The final step in the framework closes a feedback loop: interpreted insights are shared back with the household, enabling informed adjustments to routines, device use, or consent settings. This human-in-the-loop cycle ensures that the system remains assistive and privacy-respecting rather than diagnostic.

4.3.1.6 Novelty of the Layered Mapping

We propose a structured, router-centric layered mapping from packet-header features to Behavior Observation Metrics (BOM), then to DSM-5 MDD criteria via transparent fuzzy fusion and a DSM-style gate. Unlike prior strands that either remain device-centric or stop at aggregate network statistics, this formulation exposes every link, from flows to features, from features to BOM components, and from BOM to criterion likelihoods, and that in a way that also non-technical stakeholders like care-givers or parents of adolescents can inspect. The separation of *data handling* (capture, enrichment, partitioning) from *interpretation* (BOM design, memberships, weights) and *decision rules* (gate logic) makes each transformation explicit, auditable, and configurable.

- **Formulation** A router-centric, protocol-agnostic pipeline that maps packet-header features $\rightarrow$ BOM components $\rightarrow$ DSM-5 MDD criterion likelihoods using transparent membership functions, weights, and a DSM-style gate.
- **Significance** (i) Preserves privacy by operating on metadata only (no payload inspection); (ii) enables person-level attribution in multi-resident homes; (iii) yields *interpretable*, DSM-aligned indicators rather than in-transparent predictions-addressing gaps around attribution, under-used header signals, and clinical interpretability.

- **Validation Stance** The modular design supports *independent validation of each layer* (feature extraction, BOM aggregation, gate rules), so tuning or replacing one layer does not compromise the others.
- **Cross-chapter Pointers** Concrete UI, ETL, and data-flow instantiations appear in Chapter 5 (see especially Section 5.2 for the implemented architecture; Sections 5.4 and 5.4.1 for capture/partitioning and enrichment, and Section 5.5.1 for formatting and status signals). The layered mapping operationalized the Contributions stated in Section 1.8 by linking router-side observables to DSM-5–consistent indicators with transparent parameters.

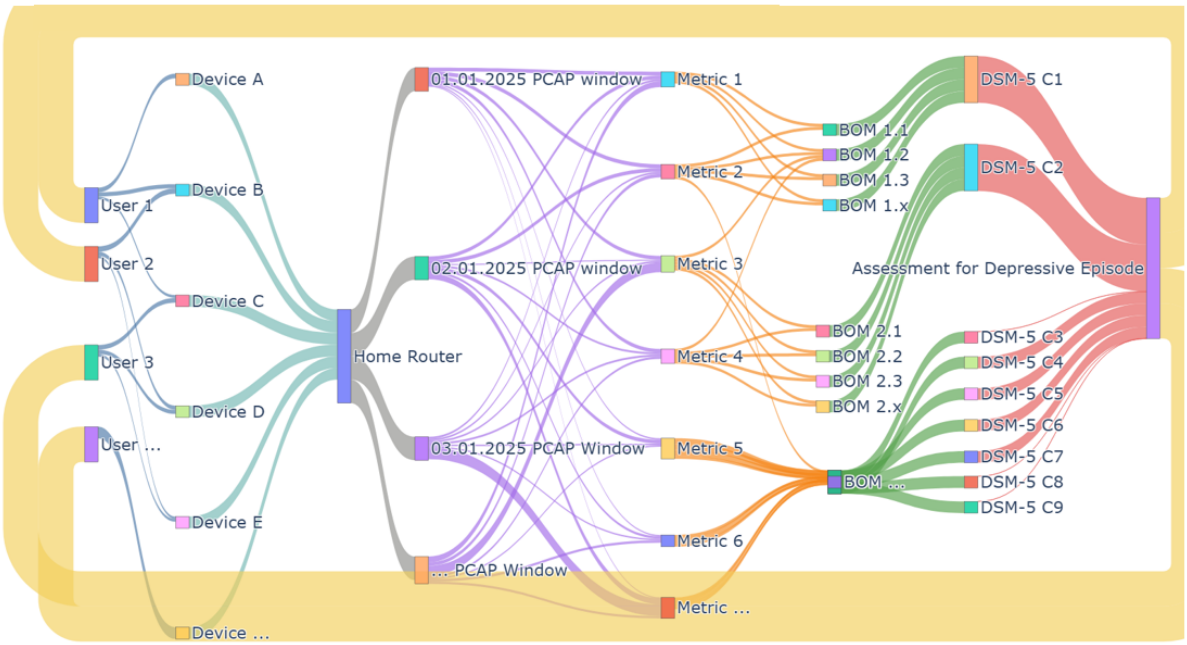


Figure 4.3: Router-centric layered mapping from household users to DSM-5 indicators with a feedback loop. From left to right: users and devices feed traffic to the home router, PCAP windows are transformed into metrics, metrics map to Behavior Observation Metrics (BOM), align with DSM-5 criteria, and an episode-level summary supports clinical assessment. The arrow on the right visualizes feedback from assessment back to the household.

Figure 4.3 summarizes our 5-layer design and the closed information-flow loop that makes router-centric observation useful over time. From left to right, household activity reaches the home router, packet captures are partitioned into time windows, windows are transformed into metrics, metrics combine into BOM and BOM align with DSM-5

indicators to support an episode-level assessment. The rightward arrow returning to the household closes the loop. When the DSM-5 MDD indicators are reviewed, the discussion may change routines or device use, which then appears in subsequent windows as rising, falling, or stable trends. Because feedback is explicit, the mapping between layers is configurable: metrics can receive higher or lower weights, thresholds can be tuned, and certain metrics can be made more sensitive to fluctuations. This iterative update keeps the system robust, improves personalization, and preserves our design intent to surface interpretable indicators rather than a diagnosis. Building on this architecture, the next Subsection 4.3.3 specifies the initial feature set we operationalize at the router, shows how these metrics map to BOM, and how BOM map to DSM-5 criteria. The selection is a pragmatic starting point that demonstrates feasibility and the range of possibilities, as also illustrated in Figure 4.4.

4.3.2 Design of Mapping

Each DSM-5 criterion (C1–C9) is supported by a small, self-contained set of criterion aligned metrics that turn router-side observables into interpretable indication signals with an explicit direction (*"pro indication"* or *"contra indication"*). Metrics are computed on temporal windows and normalized against personal baselines to preserve interpretability between individuals. This design follows the principles of privacy by design, interpretability and router-centric observability.

At the system-design level, these metrics are not ends in themselves: they are the *third layer* in a structured mapping that (i) groups related metrics into clinically meaningful BOM and (ii) maps those BOM to a DSM-5 criterion indicator. This separation of concerns, metrics $\rightarrow$ BOM $\rightarrow$ DSM-5 MDD criterion, keeps the clinical meaning visible at every step, allows profile-aware tuning via a parameter registry, and makes the pathway from quantitative network data to qualitative clinical constructs auditable. It also gives the system room to grow: additional metrics or even non-network signals can be added to a bundle without changing its clinical intent, while bundle caps prevent any single signal from dominating. Jonas Länzlinger recently pursued a conceptually similar approach by his Master's thesis [28], which focused on detecting depressive behaviors through an *audio-centered multimodal agent*. While his work used acoustic environmental signals as the primary input channel, our router-centric pipeline operates at the network layer and remains entirely content-agnostic. The two approaches are complementary: the audio-

based modality could be integrated as an additional digital environmental signal within the complementary mapping framework. This parallel structure confirms the extensibility of our design and strengthens the broader vision of multimodal, privacy-preserving behavioral analytics grounded in explainable mappings.
In our approach, we apply the network-based metric mapping consistently for all DSM-5 criteria (C1–C9). A mapping for each criterion is presented in the next Section 4.3.3. Before that catalog, we provide a brief, concrete illustrative walkthrough for a single criterion below.

4.3.2.1 Illustrative Pathway - Criterion 4 (Sleep)

To make the approach concrete, we walk through a single criterion. DSM-5's sleep criterion is mostly clinically assessed qualitative, whereas our vantage point over the router can leverage the quantitative assessment.

For C4, the BOM and their metrics (see Table B.4) are:

(BOM1) Shifted Sleep Timing We derive *digital sleep-onset time* and *digital wake-up time* from header-silence gaps, and track *onset-time variability* across days. Later-than-usual onset or large day-to-day swings indicate difficulty initiating sleep or a drifting schedule. These are direct, metadata-only correlates of what clinicians describe as delayed sleep onset or irregular sleep timing.
(BOM2) Changed Sleep Duration The *main nightly idle-gap length* serves as a privacy preserving proxy for sleep duration, compared against a personal baseline (e.g. via a deviation score). Persistent shortening suggests insomnia, persistent lengthening suggests hypersomnia. This turns a qualitative notion ("too little / too much sleep") into an auditable person-normalised signal.
(BOM3) Sleep Fragmentation Repeated, brief nocturnal reactivation appear as *micro-sessions* separated by short quiet gaps. Counting such micro-sessions and summarizing the *inter-awakening gap* captures maintenance insomnia without inspecting payloads. Clinically, this separates "can't stay asleep" from "goes to bed late."
(BOM4) Daytime Hypersomnia / Rhythm Flattening A raised *day-time idle ratio* (08–18 h) together with a higher *night/day traffic ratio* indicates napping or a flattened circadian rhythm. This bundle complements timing and duration by revealing diurnal imbalance that clinicians often infer from interviews.

In daily operation, each bundle contributes a single BOM component for a DSM-5 MDD Criterion. These components are then combined into a transparent criterion likelihood per criteria and later subjected to a DSM-style gate with duration and count rules. We defer the details and gate logic to Section 4.3.3 ff. and instantiate them in Chapter 5, with an evaluation reported in Chapter 6. A mapping of metrics→BOM→criteria for all 9 DSM-5 criteria is shown next in Section 4.3.3 and can be found in Table B.1, B.2, B.3, B.4, B.5, B.6, B.7, B.8 and B.9.

4.3.3 Mapping Procedure

Having established the layered mapping framework, our next step is to identify which measurable network features can populate these layers and ultimately inform the Behavior Observation Metrics (BOM). Feature selection in this context serves as the bridge between raw packet-level data and interpretable DSM-5 MDD indicators. It defines how much of the router-centric observation can be translated into meaningful behavioral evidence while maintaining privacy and reproducibility. Figure 4.4 summarizes this logic. The metrics extracted from the packet-header data are grouped into higher-order behaviors, which in turn form the measurable inputs to the DSM-5 MDD indicators introduced earlier. In this subsection, we outline the initial set of features implemented in our prototype, explain their rationale and describe how they are mapped to BOM and, subsequently, to DSM-5 MDD criteria. Important to mention is that this selection represents just a starting configuration that demonstrates feasibility and flexibility rather than a finalized catalog of clinical features.

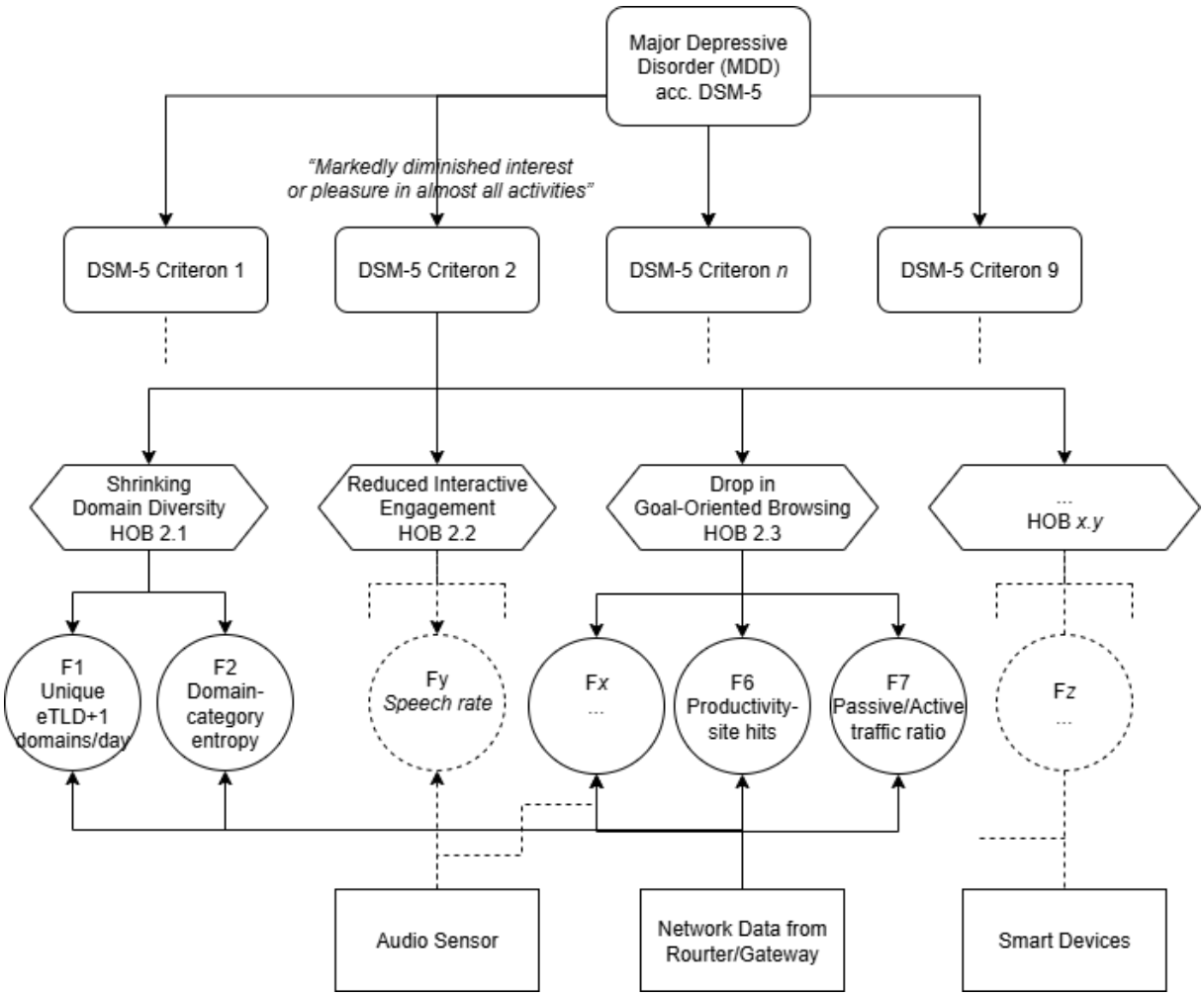


Figure 4.4: From diagnosis to indicators to behavior observation metrics and network features. The figure clarifies that we map features to higher order behaviors and then to DSM-5 MDD indicators, not to a diagnosis.

The next nine subsections introduce the metric selection and mapping for each DSM-5 MDD criterion. For each criterion you will find a short rationale, a pointer to the compact table with features and computation details and a dedicated figure that instantiates the diagram in Figure 4.4 for that single indicator.

4.3.4 Criterion 1: Depressed mood most of the day, nearly every day

Criterion 1 is a strong fit for a metadata–only pipeline. All eight features in Figure 4.5, Table B.1 can be computed from hostnames (DNS/SNI or IP/ASN when encrypted), flow timing, and byte direction. The figure in 4.5 therefore shows solid connectors for all circles. Reduced social interaction is captured by the number of distinct social sites and the typical session duration. Passive media binging is visible through long streaming sessions and a high download–to–upload ratio [11]. Rumination is approximated by low revisit diversity and short-interval repeats that surface perseverative checking [14]. A flattened rhythm is reflected by a low hourly coefficient of variation and an increasing night–to–day activity ratio [58]. None of these rely on content. They only need sessionization and time-of-day windows. This makes Criterion 1 both privacy-preserving and robust to changes, even when newer internet encryption methods like DoH or QUIC hide website names, the system can still recognize most activity patterns by looking at the servers IP address or network provider instead.

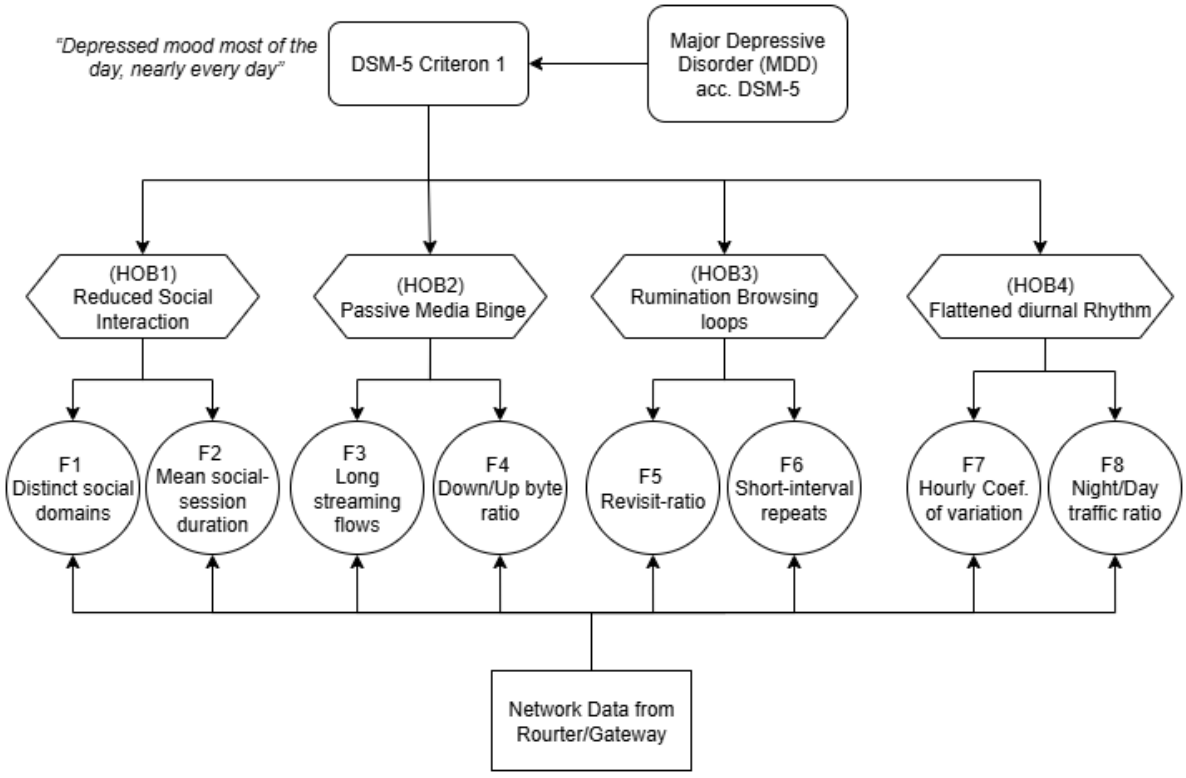


Figure 4.5: Mapping for Criterion 1. Features to behaviors to indicator. See Table B.1 for definitions.

4.3.5 Criterion 2: Markedly diminished interest or pleasure

Also the metrics used for Criterion 2 can be measured using only traffic headers and metadata-no message or page content is needed (Table B.2, Figure 4.6). A narrower set of interests shows up as visiting fewer kinds of sites ("domain diversity") and a smaller mix of topic categories ("category entropy") [23, 58]. Lower engagement looks like fewer chat sessions and a slower reply pace. We infer both from timing and tiny upload bursts, not from what was said. The passive/active balance compares heavy downloading (e.g. streaming) with short, frequent uploads (e.g., typing or clicks) to show a drift toward more passive consumption [11]. Reply latency and the passive/active ratio are heuristics, the other features are solid, where no content inspection is required.

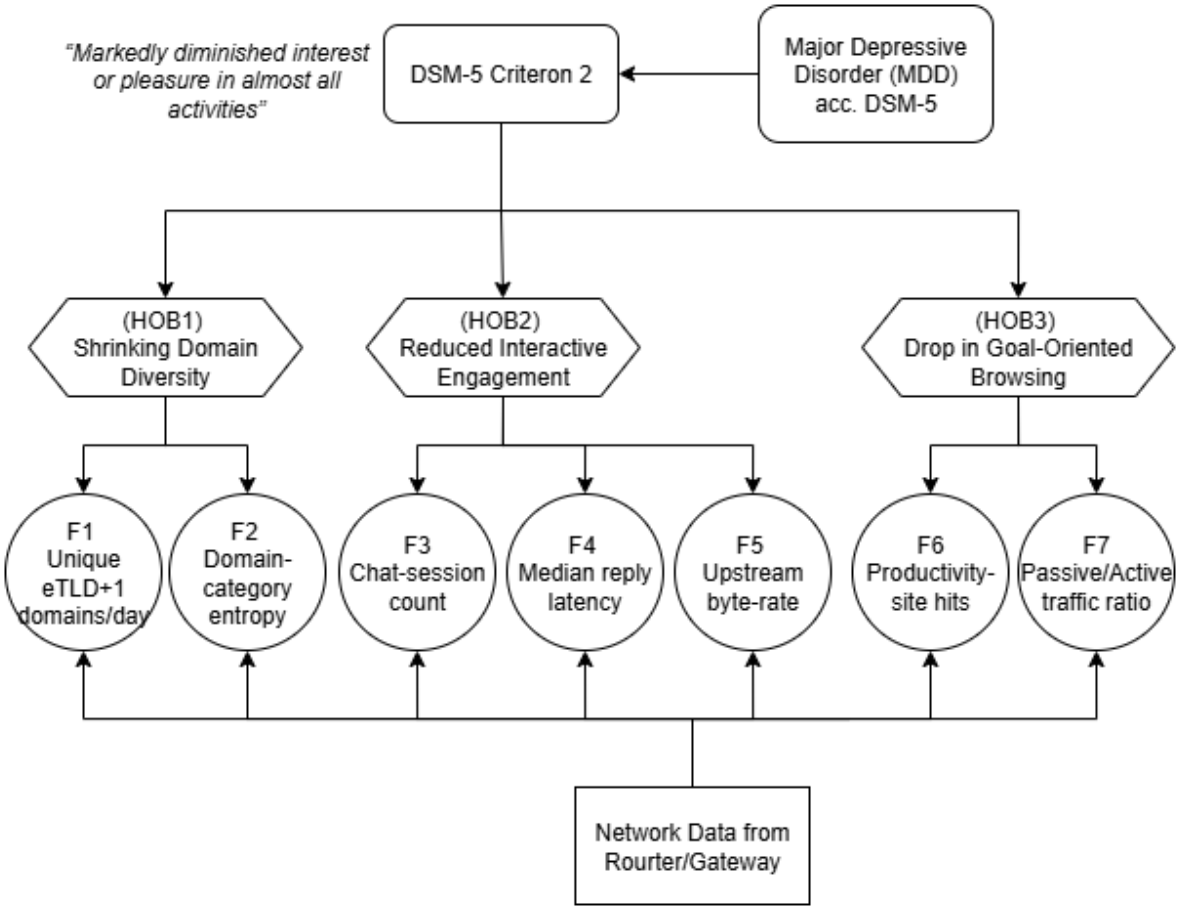


Figure 4.6: Mapping for Criterion 2. Features to behaviors to indicator. See Table B.2.

4.3.6 Criterion 3: Appetite / weight change

Criterion 3 is harder: network traffic alone only gives coarse hints (Table B.3, Figure 4.7). From headers/metadata we can see visits to delivery services and when they happen (e.g. late-night spikes), and we can spot diet/nutrition sites via broad topic labels [23]. But traffic alone cannot tell us *what was actually ordered*, *how many calories were logged*, or *whether a weight changed*. To make this criterion reliable, we should combine traffic metadata with a few consented, privacy-preserving data sources.

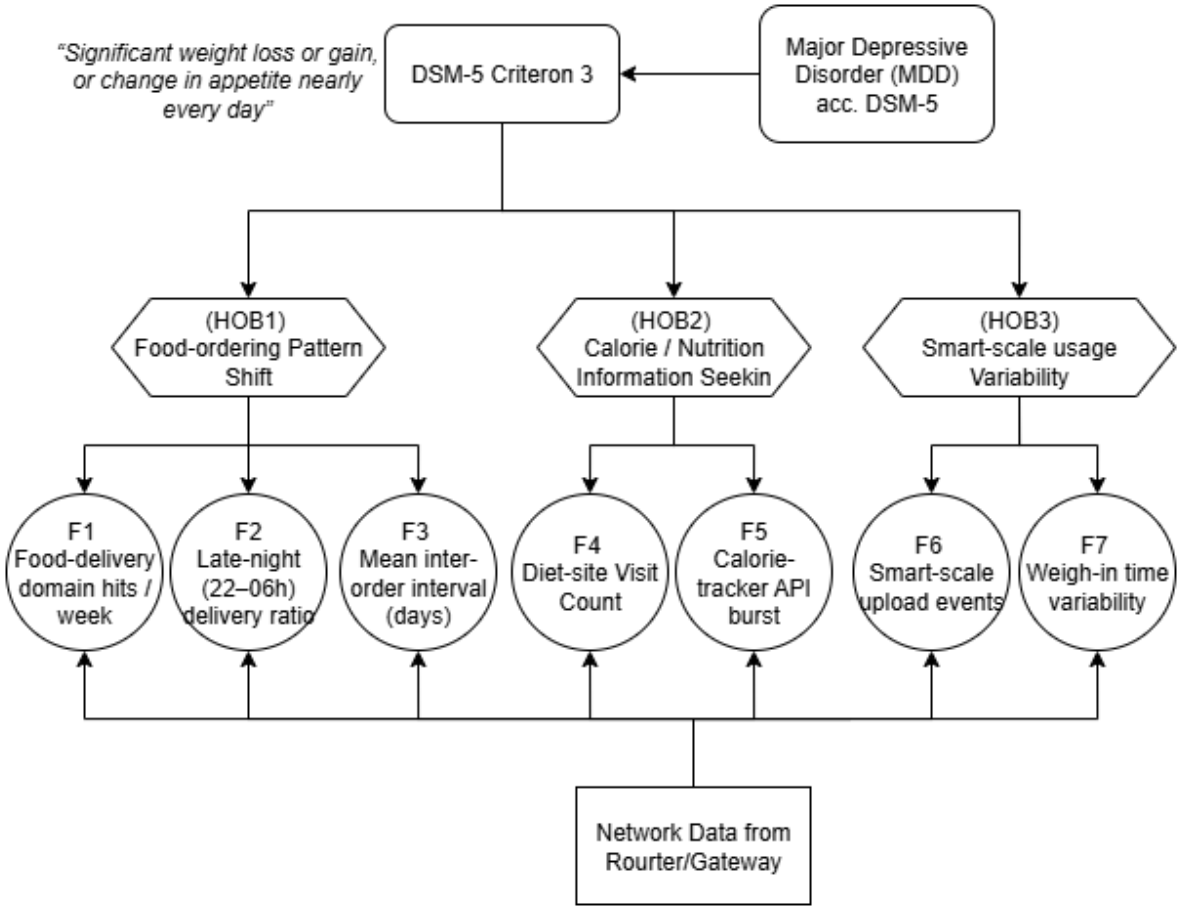


Figure 4.7: Mapping for Criterion 3. Features to behaviors to indicator. See Table B.3.

4.3.7 Criterion 4: Insomnia or hypersomnia

Criterion 4 is one of the most robust indicators of a privacy–first design. All features in Table B.4 rely on quiet periods and short bursts, not on hostnames. Digital sleep onset, wake time, and duration are inferred from long idle gaps. Fragmentation comes from counting micro-sessions at night [46, 45]. The approach agrees well with actigraphy-style ground truth in passive Wi-Fi studies [31]. Daytime hypersomnia is reflected by a high share of idle minutes and a sustained night–to–day traffic shift [19]. Figure 4.8 therefore uses only solid links. This criterion is computable from packet headers and timing metadata only.

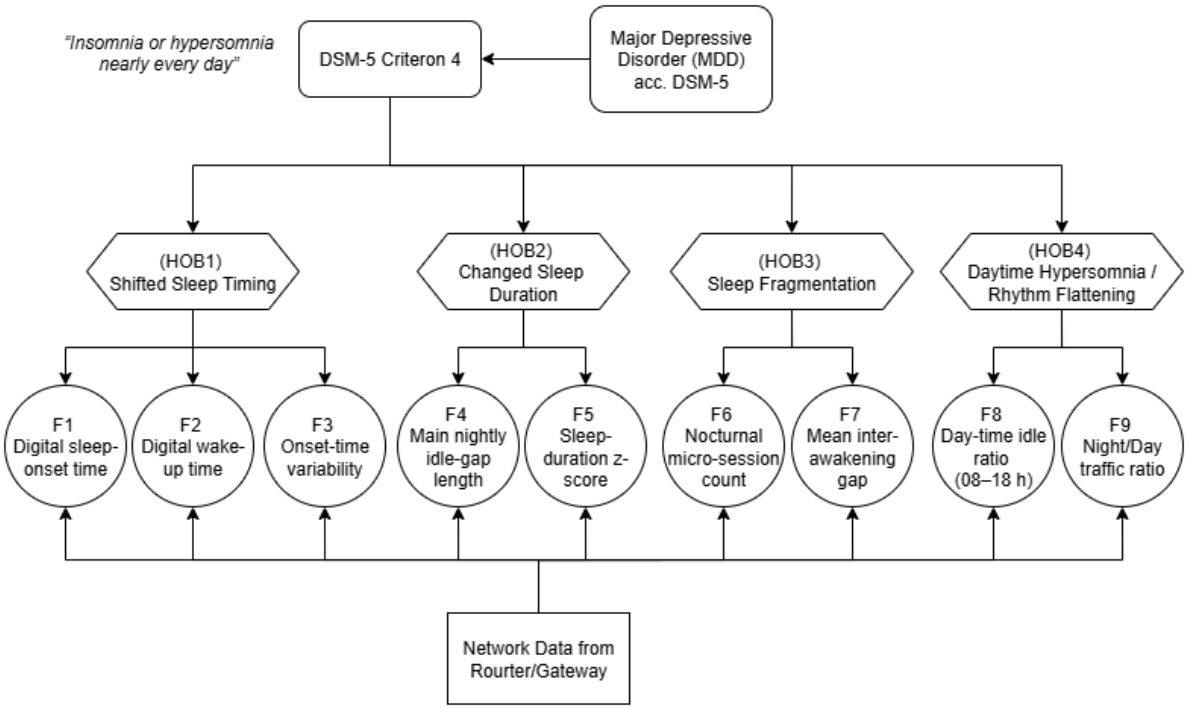


Figure 4.8: Mapping for Criterion 4. Features to behaviors to indicator. See Table B.4.

4.3.8 Criterion 5: Psychomotor agitation or retardation

For Criterion 5 we rely on router-visible events. The router records when devices connect, reconnect, or briefly drop off Wi-Fi (DHCP renewals and AP re-associations), and it shows how often very short network sessions occur. Many reconnects and many very short sessions are simple, interpretable proxies for *restlessness/agitation* [59, 46]. What the router cannot see are keystrokes or actual typing tempo. Everyday apps use encryption, so only packet headers and timing are visible—not the keys you press or their exact rhythm. While small upstream bursts during chat use might loosely correlate with how someone interacts, this is only a coarse proxy. It is far weaker than direct, consented phone- or app-level telemetry (e.g., typing dynamics) reported in prior work [64, 32]. In short, our network metrics provide privacy-preserving, coarse evidence of agitation (e.g., frequent Wi-Fi reconnects, short sessions), but they do not capture fine motor slowing/acceleration directly that would be helpful for a Criterion 5 indication. Criterion 5 is therefore hard to derive from network data alone. A stronger signal could emerge in future, optional hybrids that combine router-based burstiness with consented on-device measures (e.g., touch/scroll cadence, screen on/off rhythms, simple wearable signals), carefully balancing added sensitivity against privacy.

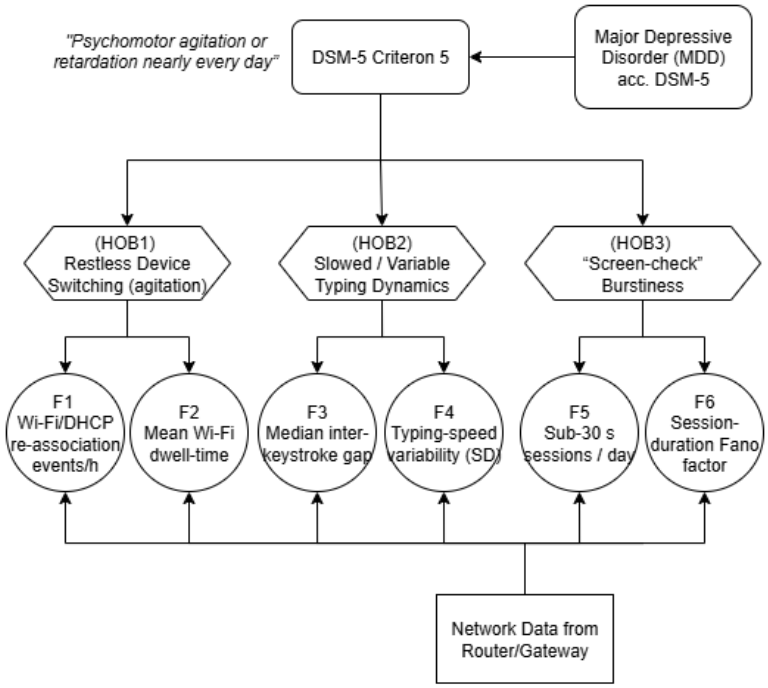


Figure 4.9: Mapping for Criterion 5. Features to behaviors to indicator. See Table B.5.

4.3.9 Criterion 6: Fatigue or loss of energy

Criterion 6 combines three metadata-friendly strands (Table B.6 and Figure 4.10). Long idle periods at midday, fewer daytime sessions and a higher idle share mark low-energy days [57, 45]. Effortless interaction decreases with fewer interactive upstream bursts and lower sent bytes per active hour [60]. The browsing tempo becomes slower and more irregular. These do not require content. They only need hourly aggregation and per-person baselines.

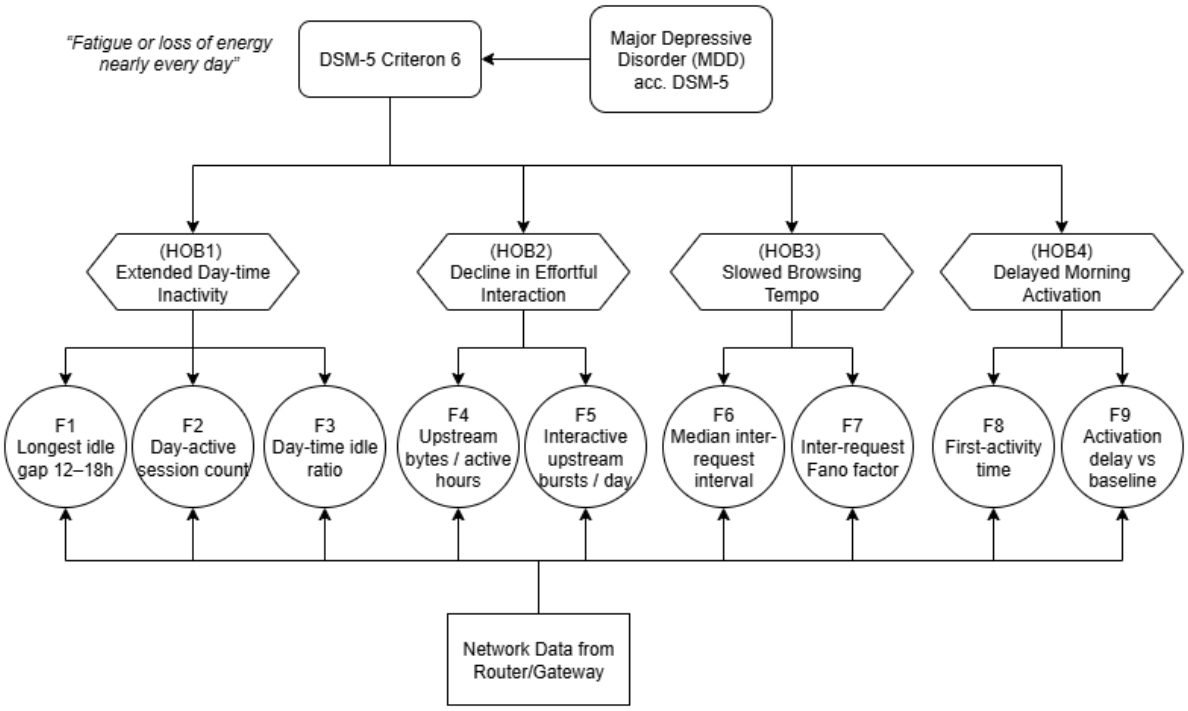


Figure 4.10: Mapping for Criterion 6. Features to behaviors to indicator. See Table B.6.

4.3.10 Criterion 7: Feelings of worthlessness or excessive guilt

Criterion 7 contains both strong domain signals and content-heavy items (Table B.7 and Figure 4.11). Visits to mental-health resources and therapist directories could be visible at the hostname level and often increase during periods of worthlessness [14]. Several items in the literature rely on path or query text, such as negative-self searches or explicit delete-account calls. These are drawn with ticked connectors to indicate that a content view would be required. The same higher order behaviors can be supported by other modalities. Voice diaries or acoustic markers could add context for self-blame, platform exports or app integrations could confirm account changes.

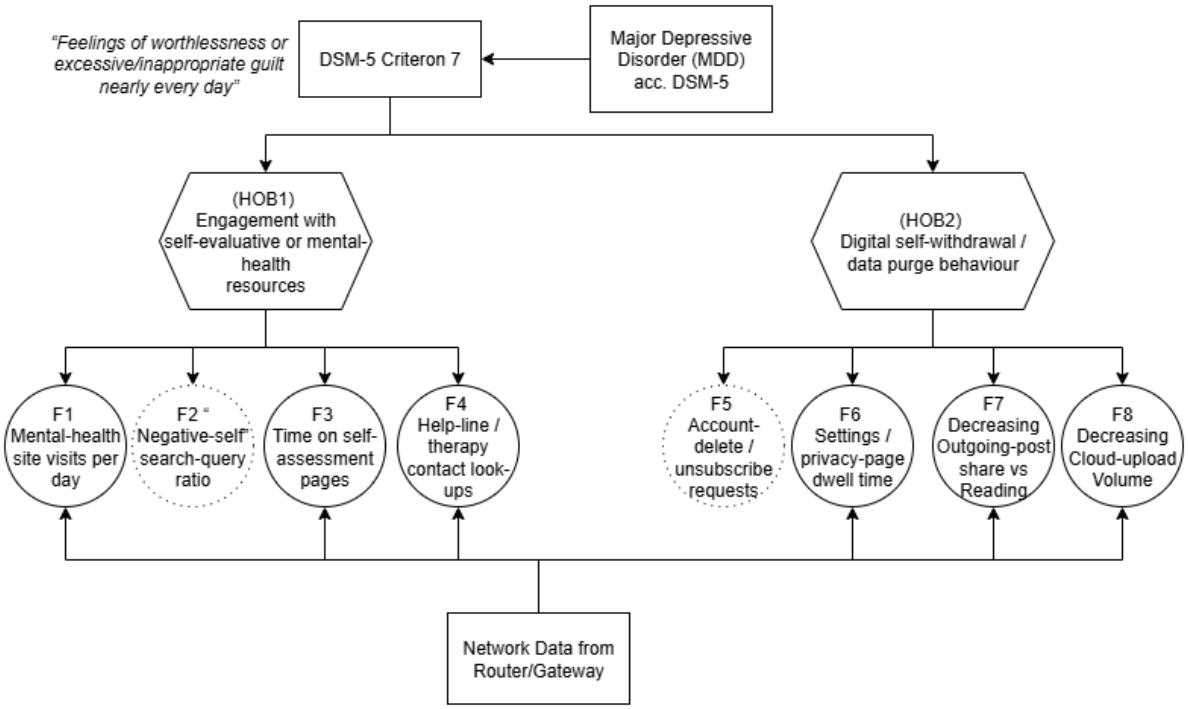


Figure 4.11: Mapping for Criterion 7. Features to behaviors to indicator. See Table B.7.

4.3.11 Criterion 8: Diminished ability to think or concentrate, or indecisiveness

The Criterion 8 can be derived from navigation timing and header-only signals (Table B.8, Figure 4.12). Fragmented focus is reflected in shorter-than-usual page dwell times, bursts of DNS lookups during tab hopping, and notification-triggered micro-sessions that interrupt ongoing activity [27, 57]. Indecisive information seeking appears as repeated returns to a search engine within a short window, a high back-navigation percentage after clicks, and a measurable time-to-first click on the results page when the destination domain changes [14]. Features that would require reading query text, such as reformulation chain length, are not computed in our privacy-preserving pipeline. The interaction tempo is approximated from small upstream bursts, which is a rough indicator rather than the truth of the keystroke. All signals are derived from packet headers, timestamps, and domain changes only, providing an interpretable, content-free view aligned with Criterion 8.

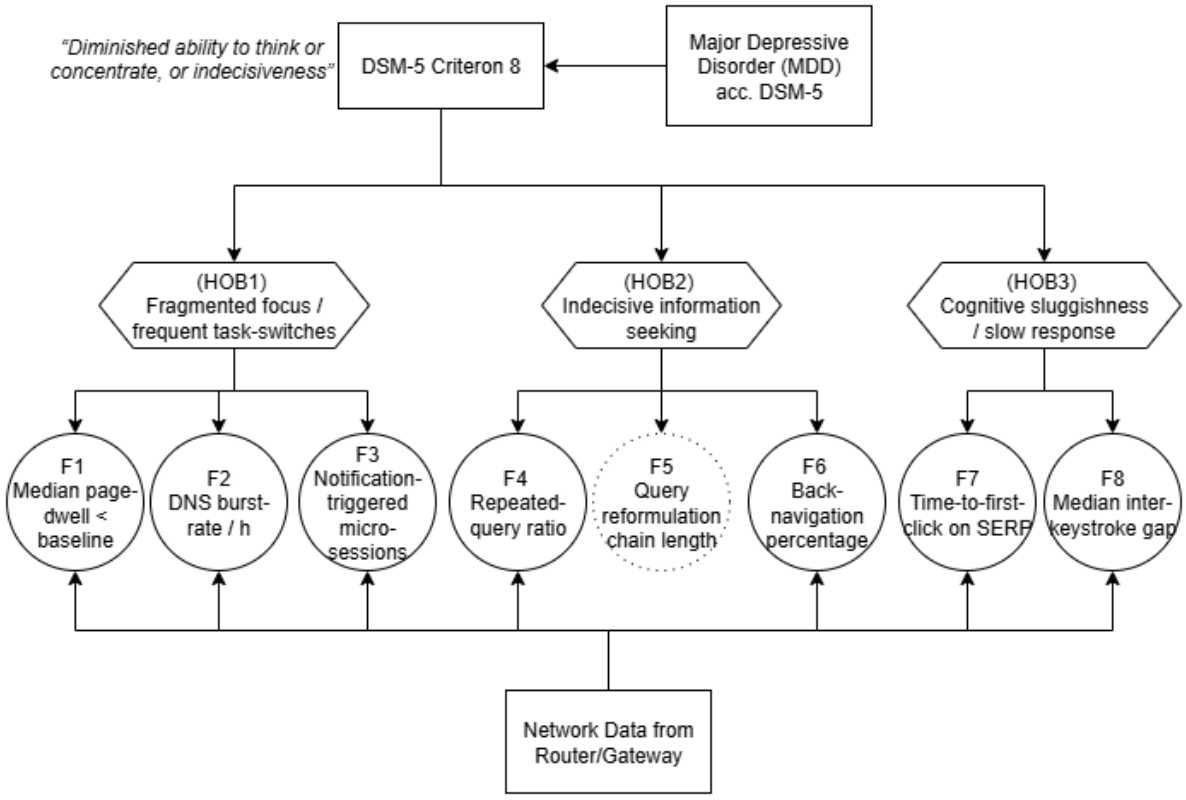


Figure 4.12: Mapping for Criterion 8. Features to behaviors to indicator. See Table B.8.

4.3.12 Criterion 9: Recurrent thoughts of death or suicidal ideation

For Criterion 9 we could rely on conservative domain-level markers shown in (Table B.9 and Figure 4.13). Crisis-line visits, therapy directories, and self-harm communities are visible without payload and have been linked to suicidal ideation (SI) trajectories [14]. Cloud-backup surges are also observable and can precede mood crashes [60]. Items that need search terms or explicit delete calls are shown with ticked lines and are not computed. Where appropriate consent exists, complementary sources such as safety-plan apps, keyboard telemetry, or clinician-managed check-ins could strengthen the mapping. Population work has shown late-night suicide-related queries [52], yet our design keeps those as non-computed placeholders to preserve privacy.

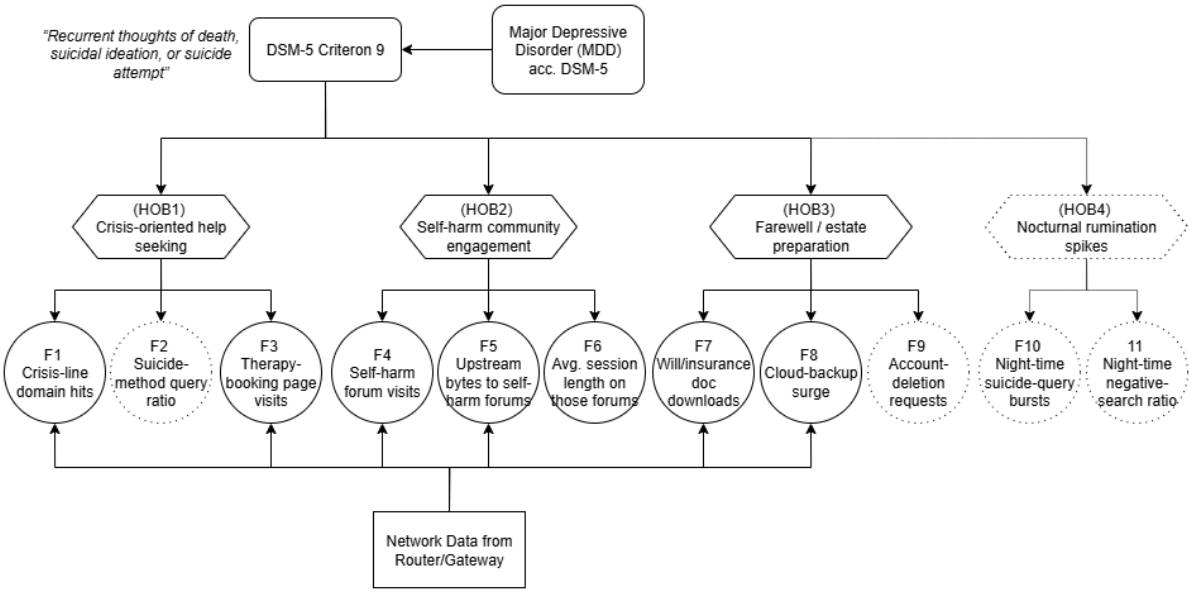


Figure 4.13: Mapping for Criterion 9. Features to behaviors to indicator. See Table B.9.

4.4 Data Flow and Computation Steps

The following section describes how the raw packets collected at the home gateway are transformed step by step into interpretable daily indicators. Each stage corresponds to a component in Figure 4.2.

4.4.1 Capture to Daily Analysis Frame

Network traffic is captured at the gateway (*Network Capture*) and organized into fixed n-minute segments within the first boundary of privacy (*Partitioned Storage*). A loader reconstructs a *daily analysis frame* by stitching the segments that fall within the selected calendar day, tolerating missing or partially written windows. During this step, records are enriched with e.g. timing fields (date, hour) and, where available, DNS/SNI hostnames and effective TLD + 1 labels to support downstream feature logic. In our system design, identity information from `Device & Session Discovery` and `Identity Resolution (IP → Profile)` can be applied to attribute events to a specific user profile in the *User Profile Registry*. When this enrichment is enabled, the resulting per-profile day frame can feed both the base record and the criterion-specific computations shown in Figure 4.2. This capability is optional and extends the MVP rather than reflecting a feature currently implemented.

4.5 Daily Likelihoods with Fuzzy Memberships and Additive Fusion

To move from individual metrics to criterion-level indicators, we introduce the concept of a *Fuzzy Additive Symptom Likelihood* (FASL). The idea of FASL is to assign each metric a fuzzy membership function that expresses how strongly its current value supports the presence of a symptom and then to combine these memberships through a transparent and weighted additive model. This approach produces a daily likelihood score per DSM-5 MDD criterion that can be interpretable, tunable and will be robust against noise in any single metric.

Before aggregation, the raw metric values of each day are *personalised* (e.g., through baseline or robust percentiles) so that "0" denotes *typical* and "1" denotes *clearly unusual* for that person. Each feature x_i is then assigned to $[0,1]$ by a simple and easily explainable membership function $\mu_i(\cdot)$ (see also alternative function in Section 4.5.2):

$$\mu_{\text{tri}}(x;\ell,m,h) = \begin{cases} 0, & x \le \ell, \\ \frac{x-\ell}{m-\ell}, & \ell < x \le m, \\ \frac{h-x}{h-m}, & m < x < h, \\ 0, & x \ge h, \end{cases} \tag{4.1}$$

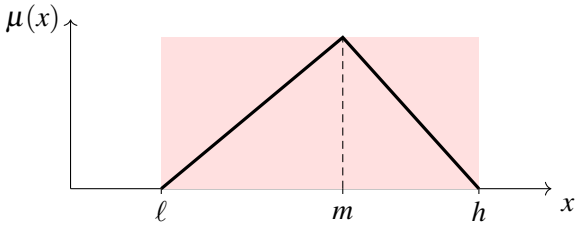
For each DSM-5 criterion k, the daily likelihood is an *additive* fusion of its fuzzy signals:

$$L_k = \underset{[0,1]}{\text{clip}} \left(\sum_{i \in \mathscr{F}_k} w_{k,i} \mu_{k,i}(x_i) \right) \tag{4.2}$$

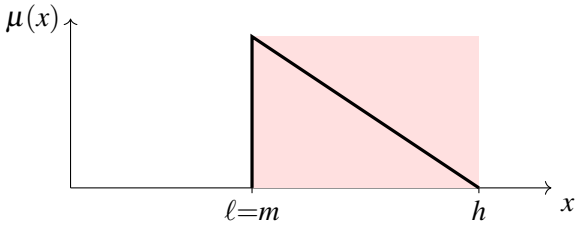
where $w_{k,i}$ are interpretable, criterion-local weights (typically normalized to $\sum_i w_{k,i} = 1$). The additive form keeps the model transparent: one can read off, for each day, which signals contributed how much to L_k.

4.5.1 Triangular Membership Function Configurations

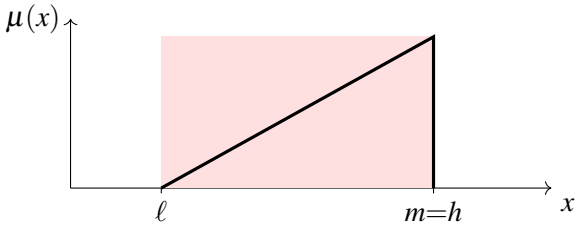
We use the triangular membership which always returns $\mu(tri) \in [0,1]$, is 0 outside $[\ell,h]$, and peaks at m. The six options below in Figure 4.14 cover the common semantics we use in practice.



Case A: $\ell < m < h$ (generic triangular membership).



Case B: $\ell = m < h$ (left-degenerate; peak at the left edge).



Case C: $\ell < m = h$ (right-degenerate; peak at the right edge).

Figure 4.14: Three geometric cases of the triangular membership function. The light-red band marks the support $[\ell, h]$ and the dashed line marks m. Labels merge automatically when endpoints coincide (e.g. ℓ=m (left-shoulder) or m=h (right-shoulder)).

4.5.1.1 Membership Example for Sleep Disturbance

Assume three signals on day d: *short sleep duration* h (shorter $\Rightarrow$ riskier), *irregular onset time* σ (14-day variability), and *late onset* m (minutes after 22:00).

(a) Sleep duration h, shorter sleep is riskier. Choose $\ell = 3.0$, $m = 6.5$, $h = 8.0$ hours. Then

$$\mu_{\text{sleep}}(h) = \begin{cases} 0, & h \geq 8.0, \\ \frac{8.0-h}{8.0-6.5}, & 6.5 < h < 8.0, \\ \frac{h-3.0}{6.5-3.0}, & 3.0 < h \leq 6.5, \\ 0, & h \leq 3.0. \end{cases}$$

This peaks near moderately short nights and returns to 0 for both very short and clearly adequate durations.

(b) Irregular onset time σ, more varriance is riskier Use $\ell = 0.2$, $m = 1.5$, $h = 3.0$ hours (circular SD over 14 days). Then

$$\mu_{\text{irreg}}(\sigma) = \begin{cases} 0, & \sigma \le 0.2, \\ \frac{\sigma - 0.2}{1.5 - 0.2}, & 0.2 < \sigma \le 1.5, \\ \frac{3.0 - \sigma}{3.0 - 1.5}, & 1.5 < \sigma < 3.0, \\ 0, & \sigma \ge 3.0. \end{cases}$$

(c) Late onset m - later onset is riskier. Use $\ell = 30$, $m = 120$, $h = 240$ minutes after 22:00. Then

$$\mu_{\text{late}}(m) = \begin{cases} 0, & m \le 30, \\ \frac{m - 30}{120 - 30}, & 30 < m \le 120, \\ \frac{240 - m}{240 - 120}, & 120 < m < 240, \\ 0, & m \ge 240. \end{cases}$$

With weights $(w_{\text{sleep}}, w_{\text{irreg}}, w_{\text{late}}) = (0.5, 0.3, 0.2)$ and observations $h = 5.8\,\text{h}$, $\sigma = 1.8\,\text{h}$, $m = 110\,\text{min}$,

$$\mu_{\text{sleep}}(5.8) = \frac{5.8 - 3.0}{6.5 - 3.0} \approx 0.80,$$

$$\mu_{\text{irreg}}(1.8) = \frac{3.0 - 1.8}{3.0 - 1.5} = 0.80,$$

$$\mu_{\text{late}}(110) = \frac{110 - 30}{120 - 30} \approx 0.889.$$

Thus the FASL daily likelihood that indicates C4 is

$$L_{\text{C4}}(d) = 0.5 \cdot 0.80 + 0.3 \cdot 0.80 + 0.2 \cdot 0.889 = 0.400 + 0.240 + 0.178 \approx 0.818.$$

This keeps the example consistent with the triangular base, ensures $\mu = 0$ outside $[\ell, h]$, and weights each contribution transparently.

4.5.2 Alternative Membership Families

The triangular membership is our default because it is interpretable, has finite support $[\ell, h]$, and is simple to tune. In some use cases, smoother or one-sided memberships are advantageous. Below we list three compatible families with compact calibration rules.

Monotone sigmoid (logistic)

$$\mu_\sigma(x; x_0, k) = \frac{1}{1 + \exp(-k(x - x_0))}, \quad k > 0. \tag{4.3}$$

$$\Delta \approx \frac{\ln 9}{k} \quad \Rightarrow \quad k \approx \frac{\ln 9}{\Delta}. \tag{4.4}$$

Use when a smooth, single-sided transition is preferred and hard cut-offs at ℓ, h are not required; for example, a gradual increase in "late onset" risk without hard thresholds. See Fig. 4.15, left.

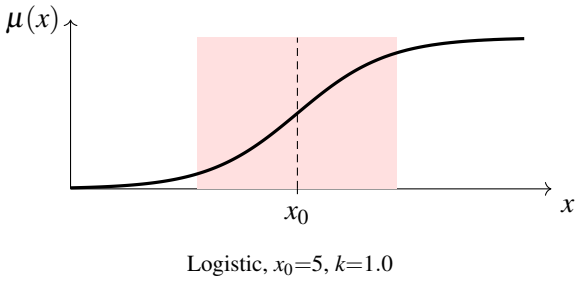


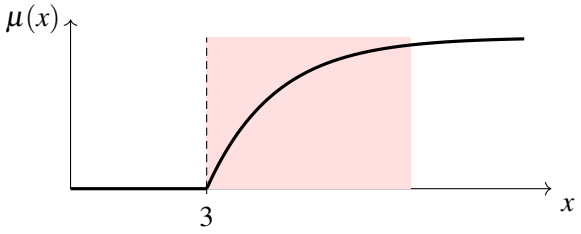
Figure 4.15: Monotone sigmoid (logistic) membership. The light-red band shows the 10–90% rise width $\Delta \approx \ln 9/k$.

One-sided exponential ramp

$$\mu_{\exp}(x; \tau, \lambda) = 1 - \exp\left(-\frac{\max\{0, x - \tau\}}{\lambda}\right), \quad \lambda > 0. \tag{4.5}$$

$$\text{choose } \lambda \text{ so that } \mu_{\exp}(\tau + \Delta) \approx 0.9 \quad \Rightarrow \quad \lambda \approx \frac{\Delta}{\ln 10}. \tag{4.6}$$

Use when evidence should accumulate progressively beyond a soft threshold τ, for example for "time-since-first-activity" after 04:00. See Fig. 4.16, middle.



Exponential, τ=3, λ=1.5

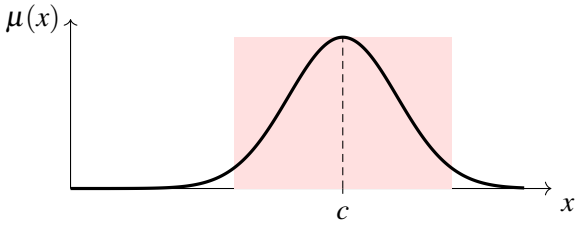
Figure 4.16: One-sided exponential ramp. The light-red band illustrates the build-up region from τ to $\tau+3\lambda$.

Unimodal Gaussian (window)

$$\mu_{\text{gau}}(x;c,\sigma) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right), \quad \sigma > 0. \tag{4.7}$$

$$\text{if } \mu_{\text{ga}}(c \pm \Delta) = \alpha \in (0,1), \quad \Rightarrow \quad \sigma = \frac{\Delta}{\sqrt{2\ln(1/\alpha)}}. \tag{4.8}$$

Use when evidence is strongest near a target c and decays symmetrically away from it, such as a "healthy" variability window for bedtimes. See Fig. 4.17, right.



Gaussian, c=6, σ=1.2

Figure 4.17: Unimodal Gaussian (window). The light-red band marks the central window $[c-2\sigma, c+2\sigma]$.

All families map $x \mapsto \mu(x) \in [0,1]$ and can replace the triangular membership in the same fusion,

$$L_k = \underset{[0,1]}{\text{clip}}\left(\sum_{i \in \mathcal{F}_k} w_{k,i}\,\mu_{k,i}(x_i)\right), \quad \sum_i w_{k,i} = 1. \tag{4.9}$$

A practical workflow for the implementation can be to start triangular for transparency, switch to sigmoid or exponential when jitter near thresholds suggests softer transitions, and use a Gaussian window when a target operating point is known.

4.5.3 Behavior Observation Metrics

In Section 4.5 we introduced the daily *Fuzzy Additive Symptom Likelihood* (FASL), which turns several personalised metrics into a single likelihood L_k for each DSM-5 MDD criterion k. In practice, it can be that multiple metrics, not only limited to metrics derived from network data, describe the same underlying observation. As an example, a "passive media binge" can be reflected by "long streaming flows" and by a high "down/up byte ratio", "flattened rhythm" can be reflected by both an hourly coefficient of variation and a night/day traffic ratio (Table B.1). If we consider and sum all metrics directly, the observations that happen to have more measurable proxies would receive more weight than observations supported by fewer proxies. This is the core limitation of an ungrouped additive fusion.

BOM components can be seen as compact, human-readable buckets that collect related metrics within a DSM-5 MDD criterion. Each component produces one bounded contribution that represents a single observation (e.g., *Reduced Social Interaction*, *Passive Media Binge*, *Rumination Browsing Loops*, *Flattened Diurnal Rhythm* for C1, Table B.1). This design has three benefits:

- It prevents single-observation dominance, ensuring that no single metric outweighs the others even when several metrics capture the same underlying behavior.
- It makes the mapping from raw metrics to DSM-5 MDD criteria transparent, so that parents, clinicians and in general non-technical users can review and discuss how each component contributes to the overall interpretation.
- It is robust to missing or sparse data, since a BOM component can still provide a contribution even if only some of its underlying metrics are available.

4.5.3.1 Aggregate Metrics and BOM

We use the same fuzzy additive synthesis with local weights at both levels of the hierarchy: first to aggregate multiple metrics into a component-level Behavior Observation Metric (BOM) slice, then to aggregate multiple BOM components into the daily criterion likelihood. Let:

$$\mathrm{FASL}_I(\{(w_\ell, z_\ell)\}_\ell) := \mathrm{clip}_I\left(\sum_\ell w_\ell z_\ell\right), \quad \sum_\ell w_\ell = 1, \tag{4.10}$$

denote an additive pooling step with normalised local weights and clipping interval I.

Metric → BOM (within-component aggregation)
For criterion k, component $b \in \mathscr{B}_k$, and metrics i assigned to b, we first fuse the metric-level membership values into a component slice. The membership functions $\mu_{k,i}(\cdot)$ and weight estimation are defined in Section 4.5:

$$E_{k,b} = \mathrm{FASL}_{[0,1]}\left(\left\{(w_{b,i}, \mu_{k,i}(x_i))\right\}_{i \in c_b}\right), \text{ optional } E_{k,b} \le c_b \le 1 \text{ (cap)} \tag{4.11}$$

BOM → Criterion (across-component aggregation)
We then fuse across components to obtain the daily likelihood series:

$$L_k = \mathrm{FASL}_{[0,1]}\left(\left\{(v_{k,b}, E_{k,b})\right\}_{b \in \mathscr{B}_k}\right), \quad \sum_b v_{k,b} = 1. \tag{4.12}$$

By design, every observation contributes at most one interpretable slice $E_{k,b}$ to L_k. The DSM-style decision rule (window M, required days N, threshold θ) is applied to the L_k series in the next stage.
Some inputs provide *indication towards* a criterion ('pro', Figure 4.18, left), while others provide *indication against* it ('contra', Figure 4.18, right). We model this bi-directionality at both levels with a signed *orientation* parameter that preserves the same FASL operator. Define a metric-level orientation $s_{k,i} \in \{+1, -1\}$ and a component-level orientation $r_{k,b} \in \{+1, -1\}$, where $+1$ denotes pro and -1 denotes contra with respect to criterion k. The signed component evidence is

$$S_{k,b} = \mathrm{clip}_{[-1,1]}\left(\sum_{i \in b} w_{b,i}\, s_{k,i}\, \mu_{k,i}(x_i)\right), \qquad \text{optionally } |S_{k,b}| \le c_b \le 1. \tag{4.13}$$

The daily criterion likelihood then pools signed component evidence (optionally with a component-level orientation $r_{k,b}$ when a whole BOM is contra):

$$L_k = \mathrm{clip}_{[0,1]}\left(\sum_{b\in\mathscr{B}_k} v_{k,b}\, r_{k,b}\, S_{k,b}\right), \qquad \sum_b v_{k,b} = 1. \tag{4.14}$$

A contra metric ($s_{k,i} = -1$) or contra BOM ($r_{k,b} = -1$) subtracts from the daily likelihood. If the weighted sum in (4.14) becomes negative on a day, clipping ensures L_k does not go below 0, which is appropriate for a likelihood-style index. The negative contribution nonetheless reduces the DSM-5 criterion indication for that day.

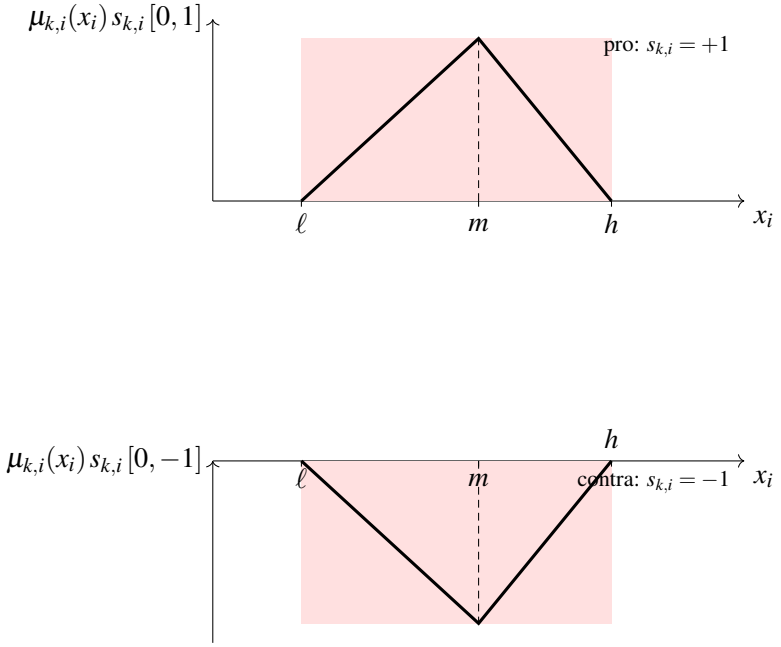


Figure 4.18: Top: indication towards (pro), $\mu_{k,i}(x_i)\, s_{k,i} \in [0,1]$ for $s_{k,i} = +1$. Bottom: indication against (contra), mirrored below the axis with $\mu_{k,i}(x_i)\, s_{k,i} \in [0,-1]$ for $s_{k,i} = -1$.

The FASL mapping and bi-directional orientations are part of the system design and are specified here to enable evaluation planning. Within implementation, described in in Chapter 5, the bi-directional metrics or BOM is not included in the current prototype due to time constraints.

Everyday Example: "Depressed mood", see also Figure 4.19
For following example we consider a relaxed rainy Sunday at home. Suppose C1 uses the four BOM components also mentioned in Table B.1:

1. *Reduced Social Interaction*: fewer distinct social domains and shorter chat sessions.
2. *Passive Media Binge*: one long streaming session and a very high down/up byte ratio.
3. *Rumination Browsing Loops*: low revisit ratio and many short-interval repeats.
4. *Flattened Diurnal Rhythm*: low hourly coefficient variation and a higher night/day traffic ratio.

Assume the personalized, normalized memberships for the day are:

Reduced Social Interaction	$\rightarrow$0.60
Passive Media Binge	$\rightarrow$0.90 and 0.80
Rumination Loops	$\rightarrow$0.50
Flattened Rhythm	$\rightarrow$0.40

A direct, ungrouped average over these multiple numbers overweight the *Passive Media Binge* signals simply because there are more of them. With BOM, we first fuse within components. For example, the *Passive Media Binge* component has two signals with fuzzy memberships 0.9 and 0.8. Each signal is multiplied by an internal weight α_i that reflects its relative importance within the component. Here we used $\alpha_1 = 0.6$ and $\alpha_2 = 0.4$, giving

$$E_{\text{binge}} = 0.9 \cdot \alpha_1 + 0.8 \cdot \alpha_2 = 0.86.$$

To prevent this single component from dominating the overall criterion, a cap $c_{\text{binge}} = 0.70$ is applied, resulting in

$$E_{\text{binge}} = \min(0.86, 0.70) = 0.70.$$

Taking equal component weights $v_{k,b} = 0.25$ across the four components (*Reduced Social Interaction*, *Passive Media Binge*, *Rumination Loops*, *Flattened Rhythm*) yields

$$L_{\text{C1}} = \tfrac{1}{4}\big(0.60 + \underline{0.70} + 0.50 + 0.40\big) = 0.55.$$

The underlined term highlights the effect of the cap: although the raw weighted average of the two features was 0.86, only 0.70 was passed forward, ensuring that the other components contribute comparably. Following example is also visualized in Figure 4.19.

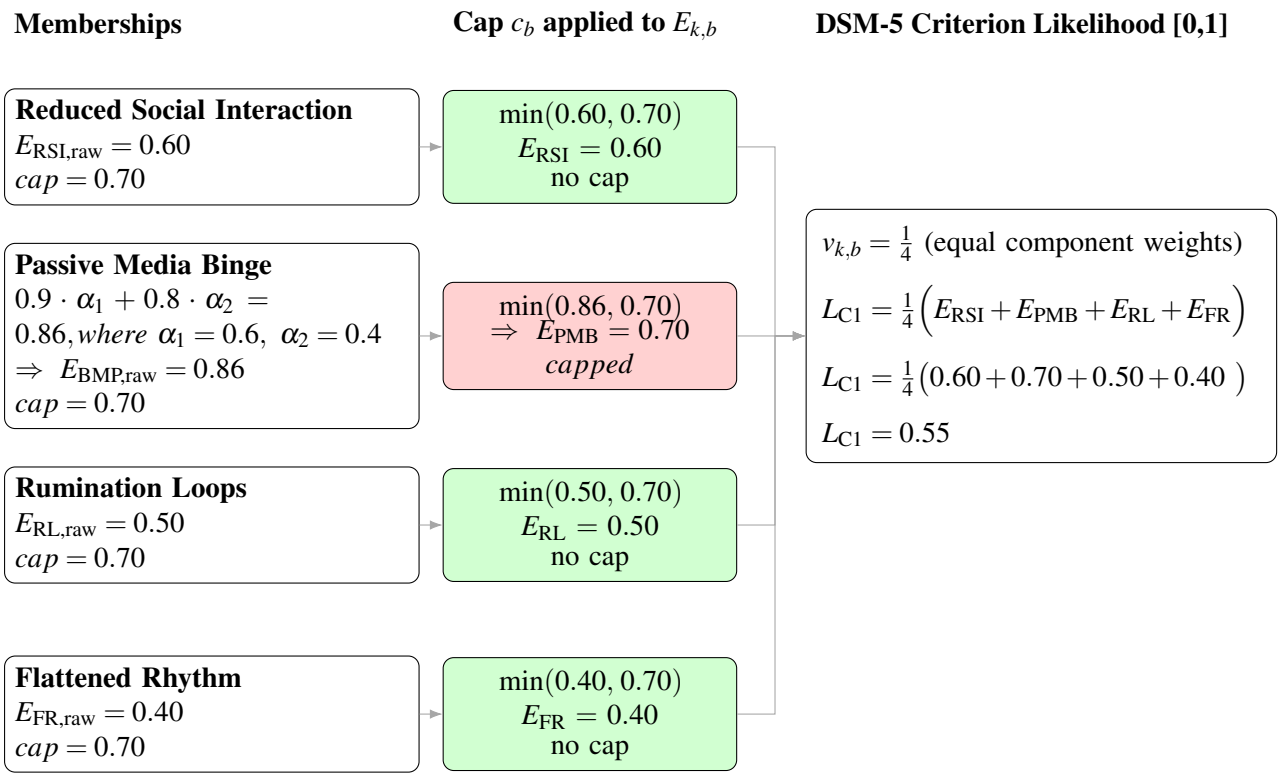


Figure 4.19: BOM Fusion Overview illustrates the overall concept of the example shown in Section 4.5.3. In this example, all BOMs are equally weighted.

The underlined cap prevents the two binge metrics from crowding out the other observations. Therefore, the result now reflects *balanced* evidence across social interaction, passivity, loops, and rhythm. In contrast, a naive average of all raw memberships here would be ≈ 0.59 and would increase even more if we added more "binge-like" metrics. With our BOM approach, we make sure it keeps the contribution per observation explicit and stable.

BOM names are plain language labels that mirror how clinicians and parents talk about behavior. In the dashboard, each criterion page should show the component values $E_{k,b}$ with an explanation of which metrics fed into them and why each falls into "OK" or "Caution" ranges that this then aligns with the conceptual objectives stated at the beginning of this Chapter 5, transparency, clinical alignment, robustness. Full metric–component mappings per-criterion are in Tables B.1–B.9, the aggregation mechanism that turns

components into daily likelihoods is in Section 4.5 and the DSM-5 gate that turns daily likelihoods into criterion-present decisions is elaborated in the next Section 4.5.4.

4.5.4 DSM-Gate and Episode Decision

The DSM-5 standard defines an MDD as the presence of ≥ 5 of 9 symptoms on *nearly every day* for at least two consecutive weeks, with at least one *core* symptom (depressed mood or loss of interest/pleasure). The DSM-Gate turns this guidance into a transparent, day-by-day decision rule: it converts continuous, interpretable daily signals into (i) a binary per-criterion presence indicator and (ii) an overall episode decision that respects the DSM-5 counting logic.

For each DSM-5 MDD criterion k we compute an interpretable daily likelihood $L_k(d) \in [0,1]$ (Section 4.5). A day counts as *positive* for this criterion if the likelihood exceeds a clinical threshold θ:

$$I_k(d) = \begin{cases} 1, & \text{if } L_k(d) \geq \theta, \\ 0, & \text{if } L_k(d) < \theta. \end{cases} \tag{4.15}$$

To operationalize the DSM-5 phrase "nearly every day", we evaluate positivity over the last M days and require at least N positive days in that window:

$$\text{present}_k(t) = \begin{cases} 1, & \text{if } \sum_{d=t-M+1}^{t} I_k(d) \geq N, \\ 0, & \text{otherwise.} \end{cases} \tag{4.16}$$

According to the DSM-5 standard, consistent settings would be $M = 14$ (two weeks) with N chosen to reflect "most days" (e.g. $N = 10$).

4.5.4.1 Counting Criteria and Core Symptom)

After checking all nine criteria, the episode decision mirrors DSM-5 MDD:

$$\text{episode}(t) = \begin{cases} 1, & \text{if } \sum_{k=1}^{9} \text{present}_k(t) \geq 5 \text{ and } \exists c \in \mathscr{C} : \text{present}_c(t) = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.17}$$

where $\mathscr{C}$ denotes the set of core symptoms (according to DSM-5 MDD standard either C1 (depressed mood), or C2 (Loss of interest or pleasure in almost all activities). Thus, an episode is *likely* when at least five criteria are present and at least one core symptom is included.

The gate parameters (M,N,θ) are auditable and configurable. Days with insufficient evidence are excluded from the window sum (data-quality rule), ensuring that decisions are not driven by missing or unreliable measurements.

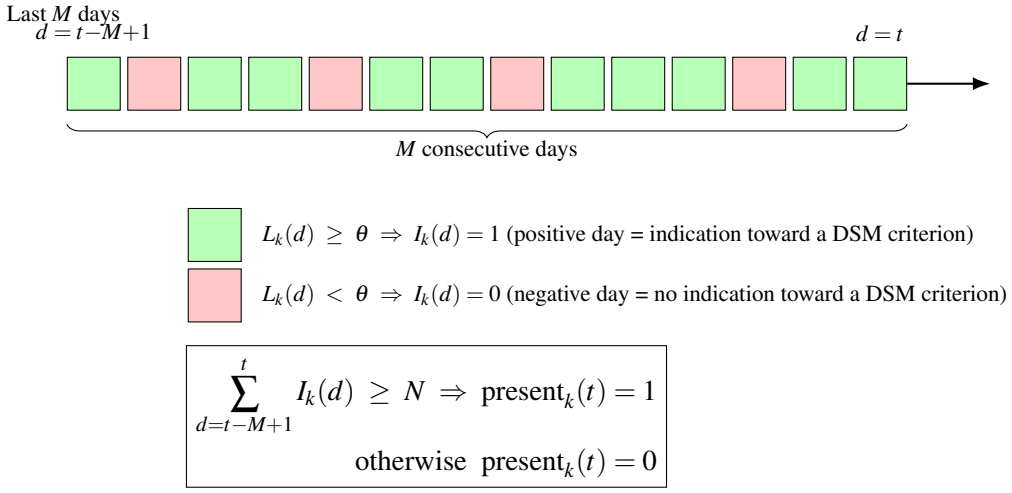


Figure 4.20: Each square represents *one day* in the last M days. A day is green if the daily likelihood for criterion k exceeds the threshold ($L_k(d) \geq \theta$), otherwise red. Count the green (positive) days and if at least N of the last M days are positive, the criterion is marked present.

4.6 Visual Guidance inside the Dashboards

The dashboards should be designed so that *non-technical users* such as parents, caregivers and clinicians-can understand, trust, and act on the indications with minimal effort. Every view mirrors the formalism introduced above: daily likelihoods $L_k(d)$, the day-level decision $I_k(d)$, the rolling gate (M,N,θ), and the DSM-consistent episode rule. The idea is that each change immediately updates the visuals, keeping the method *transparent*, *explainable*, and *reproducible* , therefore, all the parameters need to be visible and adjustable. The goal is not to teach users every metric, but to make the decision logic legible like: *what is positive today, how often in the last M days, and does that meet the DSM-style rule?*

Methodologically, the interaction follows three guiding principles:

- **Clarity over Completeness** The interface prioritizes the current per-criterion state and the minimal context needed to interpret it. Users see the day count today $L_k(d)$, the M of $\sum I_k$, and whether the DSM-style gate is met. Plain-language labels and color-safe indicators highlight the status without overwhelming with raw metrics.
- **Controllable Assumptions** Users can adjust key parameters (M, N, θ) as well as high-level aggregation settings (e.g., the contribution strength of BOM components). This supports "what-if" exploration directly in the view. Every change is logged for audit, ensuring reproducibility.
- **Provenance and Scope** Each indication is accompanied by a short *summary of the "why"* that references the components of the BOM rather than the individual signals. This makes the chain of reasoning transparent while reminding users that the results are observational and not diagnostic.

These choices are purposeful with respect to research gaps. Short excursions (G1) are visible and counted by design, rather than averaged away. The status and attributions of the profile Per-profile reduce household ambiguity (G2) and the interface states that the indications come from mostly header-derived, privacy-preserving signals (and also optional environmental sensors) to address underused but available evidence (G3) and the mapping from the signals to the BOM components to L_k and the gate outcomes replaces black-box behavior with an interpretable chain (G4). The concrete elements and layout of the user-interface that realize these principles are shown in the next Chapter 5.
In summary, this chapter presented a System Design that translates raw, privacy-compliant network observations into interpretable daily indicators that are consistent with the DSM-5 MDD. By combining feature engineering, BOM components, fuzzy aggregation, and a DSM-like gate, the system operationalizes clinical rules in a transparent and configurable manner. At the same time, the design principles for visual guidance ensure that non-technical users can interact with the results, adjust assumptions, and understand why a criterion or episode is displayed. Together, these methodological decisions close the previously identified research gaps and create a robust and explainable workflow. The following Chapter 5 describes in detail how these concepts are implemented in practice and shows the user interface and technical setup used to realize this design.

Chapter 5

Implementation

This chapter documents the implementation of my research prototype that turns raw packet captures into interpretable, privacy-preserving behavioral indicators aligned with DSM-5 MDD criteria and aggregates them into daily likelihoods. I implemented a Python-based Streamlit Inc. [51] application according to the System Design described in Chapter 4 with a PCAP–to–Parquet ETL, daily feature engineering focused on *packet headers*, a criterion-aware metric library, a *Fuzzy Additive Symptom Likelihood* (FASL), a DSM-Gate and an interactive User Interface (UI) as a reporting tool for exploration and also expert override. The System Design allows using network traffic content and richer device signals, but for this prototype we focused to header-level information (timestamps, ports, IP direction heuristics, DNS-derived hostnames).

5.1 Repository

The complete source code of our implementation is available on GitHub Repository `app_pcap_to_dsm5_public` [36]. The repository contains the Streamlit app, the metrics library, ETL utilities, configuration snapshots for FASL/DSM-Gate and all deployment assets discussed in this Chapter 5. The application is organized as a multi-page Streamlit web application (`app/pages`) with a metric library (`app/metrics`), ETL artifacts (`app/processed_parquet`), and a feature cache for fast navigation (`app/feature_cache`), in Section 5.2 the implemented architecture will be explained (see GitHub Repository Overview in the Appendix C.

5.2 Implemented Architecture

Figure 5.1 maintains the methodological layout and each component with the implemented key files and -functions. The pipeline starts with a manual upload of PCAP files into the PCAP loader module on the `User&Network Settings` page (`app/pages/03_-`

`User_and_Network_Settings.py`). Uploaded captures are partitioned into five minute Parquet segments under `app/processed_parquet/` with a deterministic suffix `__YYYYMMDD_HHMM.parquet`. This naming supports stable day and week selection.
Daily base features are computed in `app/metrics/base_features.py` using lightweight helpers from `app/metrics/common.py`, for example hostname enrichment, sessionization and direction heuristics explained in Section 5.4. The resulting all days feature table is persisted as a compact cache in `app/feature_cache/*.csv`. Keeping this cache on "disk" keeps the pipeline portable and simple to rerun in local and containerized environments.
per-criterion computations are implemented in `app/metrics/criterion1..9.py`. Each module exposes `CriterionX.compute()` and uses the resolver registry in `app/metrics/common.py` via `resolve_value()` and `FUNC_REGISTRY`. This design allows declarative metric definitions that remain transparent and reproducible.
Two Streamlit pages consume those metrics. The Network Metrics page (`app/pages/02_-Network_Metrics.py`) provides daily drill downs and comparisons over time. Metric selection is explicit and manual with a multi select per DSM 5 tab. The Early Signs Overview (`app/pages/01_Early_Signs_Overview.py`) computes the Fuzzy Additive Symptom Likelihood and applies the DSM-5 Gate. Weights and triangular membership functions are read from versioned JSON presets in `app/utils/fasl_config_*.json`. Optionally `app/utils/auto_tune.py` proposes parameter sets from historical distributions. Accepted values are written back to JSON and applied at runtime. Some artifacts are present but not wired into calculations. The User Profile Registry (`app/user_-profiles.json`) and device discovery or identity mapping are decoupled and therefore have no arrow into the metric pipeline in Figure 5.1.

This implementation is an MVP. To keep the system simple and portable, we intentionally avoided the time-series database, message queuing, and stream processors. Instead, we rely on simple and transparent persistence tools to be able to make quick changes if needed. In general, our implementation favors clarity and control. Deterministic partitions, a stable daily feature record, explicit per-criterion computations and a JSON driven FASL and gate step make the system easy to audit, reproduce, and evolve while keeping the dashboards responsive and interpretable.

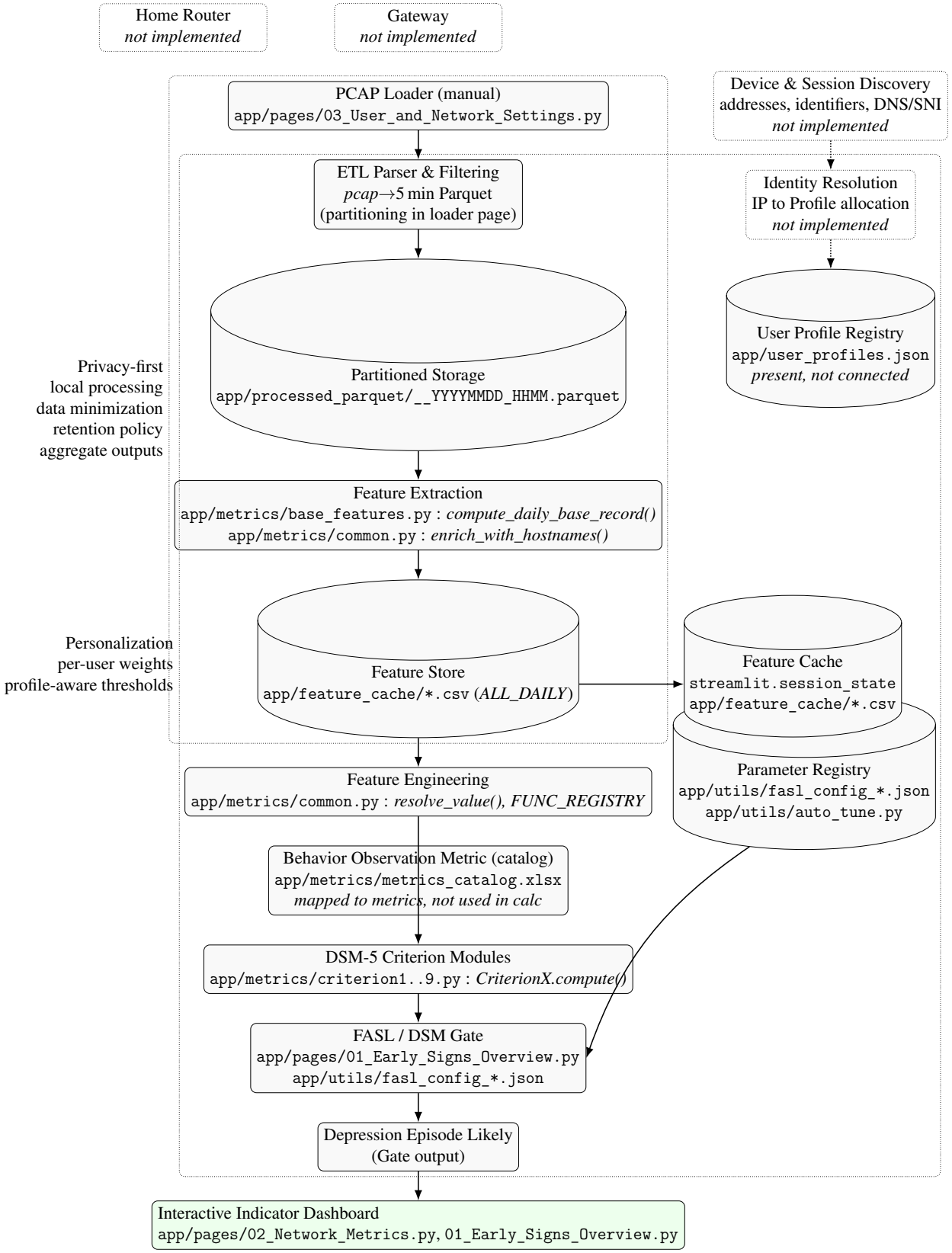


Figure 5.1: MVP Process Flow / Architecture Overview

5.3 Network Traffic Data Sources

In order to exercise the router centric pipeline end to end as drawn in Figure 5.1, we combined two complementary public datasets. The first provides heterogeneous and use case rich traffic that we used to validate that features built from headers and timing separate behavioral classes. The second provides longitudinal continuity across many consecutive days that we used to establish rolling baselines and to evaluate the gate logic. For heterogeneity and realistic usage situations we relied on the Real Network Environment dataset by Jiang and colleagues which publishes labeled traces captured from two vantage points named BRAS and ONU [25]. We used scenario labels such as `game`, `streaming` and `instant message` to sanity check that shape level signals are discriminative without inspecting payload. Long unidirectional streaming sessions, bursts of DNS requests and chat style turn taking were recognizable from timing and sizes alone. Our metric implementations are for now based on timestamped headers, directions, byte counts and SLD categories.

For continuity we used the teaching materials curated by Erik Hjelmvik for the FIRST workshop series [20]. The material provides a timeline of approximately 40 consecutive days with realistic background activity and staged events. We ingest daily logs and treat them as one longitudinal series that supports a random baseline building weekday pattern analysis and gate evaluation over windows of M days. After parsing the PCAP files we resolve destination IPs to hostnames and derive eTLD+1 domains. This yields tables with Timestamp, Source IP, Destination IP, Length, Protocol and Ports, and SLD. From these we compute daily base features and criterion specific features.

Together, the properties of these two datasets could support the claim that a router centric and payload-free pipeline can generate interpretable daily likelihoods for DSM-5 MDD criteria, this will also be discussed in Chapter 6. Both datasets are public and used here strictly for research. The material contains realistic background traffic but does not represent any participants personal ground truth; therefore, the MVP emphasizes that the underlying data do not represent actual user behavior and must not be used for diagnosis.

5.4 PCAP Ingestion and 5-minute Partitioning

To capture network traffic over, e.g. a gateway, we build a *PCAP Loader* that can process one or more PCAP/PCAPNG files and convert them into uniform, time-indexed fragments that the rest of the pipeline can consume. Each capture is partitioned into contiguous five-

minute Parquet files under app/processed_parquet/<dataset>, Appendix C.1, A9), using the pattern <dataset>__YYYYMMDD_HHMM.parquet. During ingestion we perform only light, header-preserving normalization like the timestamps are parsed to a common granularity and coerced to datetime (app/pages/01_Early_Signs_Overview.py, Appendix C.1, A2), minimal host reduction to registrable domains (eTLD+1) is done with small helpers in app/metrics/common.py, Appendix C.1, A5) and a five-minute idle rule is used later for sessionization via sessions_from_timestamps (Listing 5.1) in the same module.

Listing 5.1: Sessionization helper (app/metrics/common.py)

```
5.1.1 def sessions_from_timestamps(df: pd.DataFrame, gap_sec: int = 300)
      -> List[Tuple[pd.Timestamp, pd.Timestamp]]:
5.1.2     """Split sorted timestamps into sessions separated by > gap_sec
      seconds."""
5.1.3     if df.empty or "Timestamp" not in df.columns:
5.1.4         return []
5.1.5     times = sorted(pd.to_datetime(df["Timestamp"]).tolist())
5.1.6     if not times:
5.1.7         return []
5.1.8     sessions: List[Tuple[pd.Timestamp, pd.Timestamp]] = []
5.1.9     start = times[0]
5.1.10    for i in range(1, len(times)):
5.1.11        if (times[i] - times[i-1]).total_seconds() > gap_sec:
5.1.12            sessions.append((start, times[i-1]))
5.1.13            start = times[i]
5.1.14    sessions.append((start, times[-1]))
5.1.15    return sessions
```

To support interpretation without requiring payload, the settings page (Section 5.7) surfaces a merged catalog of *observed* local IPs across the processed Parquet files so that addresses can be mapped to user profiles. This implementation, including the UI components, can be found in app/pages/03_User_and_Network_Settings.py (Appendix C.1, A4). This mapping will feeds direction heuristics and can be used by several metrics (e.g., upstream/downstream shares, social out-going ratios).

5.4.1 DNS and Domain Enrichment

After the five-minute Parquet partitioning (Section 5.4), each partition is passed through a light resolver that reconstructs a per-partition mapping from resolved IP addresses to hostnames and then reduces those hostnames to registrable domains (eTLD+1). The resolver prefers the DNS answer name (`DNS_ANS_NAME`) and returns to the query name (`DNS_QNAME`) when the answers are unavailable. Every IP listed in `DNS_ANS_IPS` is associated with the chosen name. A destination host is then assigned to each flow by looking up its `Destination IP`. When no mapping is found, the original DNS fields are retained. Finally, a tiny helper performs a naïve eTLD+1 reduction by taking the last two labels of the host (for example, `mail.google.com` → `google.com`), which is sufficient for the SLD-level features and curated domain sets used later in the pipeline (`app/metrics/common.py`). The end result is that every row has a `ResolvedHost` and a `SLD` column that downstream feature builders consume (`app/metrics/common.py`). The conceptual implementation is the following.

Listing 5.2: DNS Enrichment, simplified, `app/metrics/common.py`.

```
5.2.1 def build_dns_map(df: pd.DataFrame) -> dict[str,str]:
5.2.2     """Concept-only: map IPs -> hostnames from DNS rows."""
5.2.3     ip2host={}
5.2.4     for _, r in df[df["IsDNS"]==True].iterrows():
5.2.5         # Prefer DNS_ANS_NAME; fallback to DNS_QNAME
5.2.6         name = r["DNS_ANS_NAME"] if ("DNS_ANS_NAME" in r and r["
          DNS_ANS_NAME"]) else r["DNS_QNAME"]
5.2.7         # Assume r["DNS_ANS_IPS"] iterable (e.g.,
          ["1.2.3.4","5.6.7.8"])
5.2.8         for ip in r["DNS_ANS_IPS"]:
5.2.9             ip2host[ip] = name
5.2.10    return ip2host
```

This enrichment step is intentionally lightweight to keep memory bounded at the five-minute granularity and to avoid external dependencies; it delivers exactly the SLD-level signals that multiple day-level features require, such as the unique-domain count (`C2_F1_UniqueSLD`) and category hits built on curated SLD lists.

Listing 5.3: Hostname enrichment, SLD, simplified, app/metrics/common.py.

```
5.3.1 def enrich_with_hostnames(df: pd.DataFrame) -> pd.DataFrame:
5.3.2     """Concept-only: attach ResolvedHost and SLD."""
5.3.3     out = df.copy()
5.3.4     ip2host = build_dns_map(out)
5.3.5     out["ResolvedHost"] = out["Destination IP"].map(ip2host)
5.3.6     # Fill gaps from DNS names on the rows themselves
5.3.7     out.loc[out["ResolvedHost"].isna(),"ResolvedHost"] = out["
          DNS_ANS_NAME"].fillna(out["DNS_QNAME"])
5.3.8     out["SLD"] = out["ResolvedHost"].apply(_sld)
5.3.9     return out
```

5.4.2 Day-level base Records

Once the five-minute partitions are enriched with hostnames and eTLD+1 domains, the pipeline aggregates them into a single day-level record that acts as the backbone for all higher-order analyses.

The function `compute_daily_base_record` (in `app/metrics/base_features.py`) consolidates raw traffic events into a compact schema that is stable across datasets and resilient to missing fields. Each output record corresponds to one observation day and contains a fixed set of descriptive statistics-such as packet counts, timing gaps, activity ratios, and byte directionality-that together provide a coarse but robust summary of digital behavior.

At runtime, the function proceeds through a consistent sequence of transformations. It first harmonizes timestamps, orders packets chronologically, and extracts the hour of day to support time-of-day features. It then calculates several temporal and rhythm descriptors from the per-minute activity series. These include the fraction of traffic occurring between midnight and 05:00 (`LateNightShare`), the maximum observed inactivity gap (`LongestInactivityHours`), and the ratio of active minutes at night versus day (`ND_Ratio`). Together, these quantities capture the broad structure of a user's daily routine, offering simple yet interpretable proxies for sleep onset, continuity, and circadian regularity.

Beyond timing, the record also maintains direction-aware byte counters (`down_bytes_today`, `up_bytes_today`) that estimate inbound versus outbound activity without deep packet inspection. This separation enables features such as upload/download asymmetry, passive-active ratios, and effortful interaction scores that recur in multiple criteria. Social

and messaging domains are detected using curated lists (e.g., SOCIAL_SLDS, MSG_PORTS) to compute social-engagement baselines such as distinct social domains and mean session duration, both reused in C1 and C2 metrics.

Listing 5.4: Selected base features (app/metrics/base_features.py)

```
5.4.1 #  TODAY: LateNightShare (0005)
5.4.2 if total_packets > 0:
5.4.3     night_packets = int(d["Hour"].isin([0, 1, 2, 3, 4, 5]).sum())
5.4.4     rec["LateNightShare"] = night_packets / float(total_packets)
5.4.5 else:
5.4.6     rec["LateNightShare"] = float("nan")
5.4.7
5.4.8 # -- TODAY: LongestInactivityHours
5.4.9 if total_packets >= 2:
5.4.10     diffs = np.diff(d["Timestamp"].values).astype("timedelta64[s]").
          astype(int)
5.4.11     rec["LongestInactivityHours"] = float(diffs.max() / 3600.0) if
          diffs.size else float("nan")
5.4.12 else:
5.4.13     rec["LongestInactivityHours"] = float("nan")
5.4.14
5.4.15 # -- TODAY: per-minute activity + rhythm metrics
5.4.16 per_min = d.set_index("Timestamp").assign(cnt=1)["cnt"].resample("1
          Min").sum()
5.4.17 active_min = (per_min > 0).astype(int)
5.4.18 ...
5.4.19 rec["ActiveNightMinutes"] = float(night_active)
5.4.20 rec["ActiveDayMinutes"] = float(day_active)
5.4.21 rec["ND_Ratio"] = (night_active / float(day_active)) if day_active >
           0 else float("nan")
```

5.5 DSM-5 Criterion Modules

Each DSM-5 criterion is implemented in a dedicated module app/metrics/criterion*.py, Appendix C.1, A7. The pattern is consistent: a small schema for human-readable labels, formatting, and status rules, plus a compute method that resolves values either from the day record or from specialized functions (e.g., DNS

bursts). All implemented metrics can be found under `app/metrics/criterion*.py`, Appendix C.1, A7.

5.5.1 Formatting and Status

A central goal of the design mentioned in Chapter 4 is that each metric is human-readable, comparable across days and transparently classified without hidden heuristics. To achieve this, I introduced a small, uniform *status and formatting layer* in `app/metrics/base.py` that every metric can use before rendering.

- **Single explicit Rule** We implemented a scalar *OK threshold* and a *direction* flag (`higher_is_worse`) and we avoided opaque percentile bands at this stage so that analysts can see and reason about the exact cut-off used (the value appears in the UI text and in the details dialog).
- **Graceful handling of Missing Values** When a value is missing, non-finite, or not applicable for a given day, the status is `N/A`. This prevents accidental "green by default" and makes data gaps visible,
- **Stable Ordering and Coloring** A global `STATUS_ORDER` map ensures that tiles and tables sort consistently (`OK` $\prec$ `Caution` $\prec$ `N/A`) and use the same color palette across pages. This consistency matters when scanning many metrics quickly.
- **Separation of Concerns** The computation produces a *payload* (via `make_metric`) that contains raw value, formatted text, status tuple, and optional LaTeX snippets. The UI components should only render and not embed business rules.

For each metric the criterion module computes a numeric value, then calls `status_from_thresholds(value, ok, higher_is_worse)` to obtain the label/color pair. The human-readable range explanation is generated via `thresholds_text(...)` and shown beneath the tile or inside the details dialog. Finally, `make_metric(...)` assembles a standard dictionary that the UI consumes. The optional `range_cfg` field preserves the exact parameters (`{"ok": ..., "higher_is_worse": ...}`) so analysts can audit, reproduce or adjust them later (Section 5.6 where these parameters align with the membership orientation).

This minimal scheme fits the header-only scope (Section 5.3) and keeps decisions explainable: a reader can reproduce `OK`/`Caution` with a pencil from the displayed threshold and direction. At the same time, it integrates cleanly with the FASL memberships in

Section 5.6: when `higher_is_worse` is `True`, the triangular membership is oriented accordingly; when it is `False`, the membership is inverted. Thus, the qualitative UI status and the quantitative aggregation remain coherent.

Listing 5.5: Threshold helpers (`app/metrics/base.py`)

```
5.5.1 STATUS_ORDER = {"OK": 0, "Caution": 1, "N/A": 2}
5.5.2
5.5.3 def status_from_thresholds(value, ok, warn=None, higher_is_worse=
      True):
5.5.4     """
5.5.5     Three Conditions: OK, Caution, N/A.
5.5.6     - higher_is_worse=True: value <= ok -> OK, else Caution
5.5.7     - higher_is_worse=False: value >= ok -> OK, else Caution
5.5.8     """
5.5.9     if value is None or (isinstance(value, float) and np.isnan(value
      )):
5.5.10        return ("N/A", "blue")
5.5.11    if higher_is_worse:
5.5.12        return ("OK", "green") if value <= ok else ("Caution", "
      orange")
5.5.13    else:
5.5.14        return ("OK", "green") if value >= ok else ("Caution", "
      orange")
5.5.15
5.5.16 def thresholds_text(ok, warn=None, higher_is_worse=True):
5.5.17    if higher_is_worse:
5.5.18        return f"Status ranges: **OK <= {ok}**; **Caution > {ok}**
      *(higher is worse)*"
5.5.19    else:
5.5.20        return f"Status ranges: **OK >= {ok}**; **Caution < {ok}**
      *(lower is worse)*"
```

5.5.2 Domain Labeling and Traffic Masks

Criterion modules import a small set of reusable labeling primitives, namely curated domain sets (e.g., `SOCIAL_SLDS`, `PRODUCTIVITY_SLDS`) and light, composable masks (e.g., `chat_mask`, `streaming_inbound_mask`). These helpers allow each criterion

implementation to select semantically meaningful traffic with one line of code while keeping payload untouched and privacy intact.

5.5.2.1 Curated eTLD+1 Label Sets

After DNS enrichment (Section 5.4.1), each row carries an `SLD` (registrable domain). We maintain small, auditable lists that map these `SLD`s to coarse categories, for example `SOCIAL_SLDS`, `STREAMING_SLDS`, `PRODUCTIVITY_SLDS`, `FOOD_DELIVERY_SLDS`, and others (`app/metrics/common.py`). Using these lists, selection becomes an interpretable set-membership test, instead of fragile regex rules or payload parsing.

5.5.2.2 Masks for Selection and Direction

Two tiny direction predicates identify *public→private* (inbound) and *private→public* (outbound) flows via the RFC1918 checks over the `ipaddress` IPv4/IPv6 manipulation library [41]. Designed on top, `chat_mask` and `streaming_inbound_mask` use a union of category membership and ports to isolate messaging and streaming traffic. The same masks are reused across criteria when computing session counts, reply latencies, passive/active ratios, or upstream byte shares.

Listing 5.6: Direction + chat selection (`app/metrics/common.py`)

```
5.6.1 def is_private_ip(ip_str: str) -> bool:
5.6.2     try: return ipaddress.ip_address(ip_str).is_private
5.6.3     except Exception: return False
5.6.4 def is_outbound(src: str, dst: str) -> bool:
5.6.5     return is_private_ip(src) and (not is_private_ip(dst))
5.6.6 def is_inbound(src: str, dst: str) -> bool:
5.6.7     return (not is_private_ip(src)) and is_private_ip(dst)
5.6.8 MSG_PORTS = [5222, 5223, 443]  # XMPP / TLS (most chats multiplex
        on 443)
5.6.9 def chat_mask(df: pd.DataFrame) -> pd.Series:
5.6.10     mask = pd.Series(False, index=df.index)
5.6.11     if "SLD" in df.columns: mask |= df["SLD"].isin(SOCIAL_SLDS)
5.6.12     if "Destination Port" in df.columns: mask |= df["Destination
       Port"].isin(MSG_PORTS)
5.6.13     return mask
```

The streaming selector composes category membership (`STREAMING_SLDS`) with the inbound direction, so it only captures *consumption* flows from public hosts to private devices.

Listing 5.7: Inbound streaming selector (app/metrics/common.py)

```
5.7.1 def streaming_inbound_mask(df: pd.DataFrame) -> pd.Series:
5.7.2     mask = pd.Series(False, index=df.index)
5.7.3     if "SLD" in df.columns: mask |= df["SLD"].isin(STREAMING_SLDS)
5.7.4     if {"Source IP","Destination IP"}.issubset(df.columns):
5.7.5         mask &= df.apply(lambda r: is_inbound(r["Source IP"], r["
          Destination IP"]), axis=1)
5.7.6     return mask
```

5.5.3 C1 - Sleep Disturbance

One metric that indicates towards Sleep disturbance (C1) mixes circadian regularity/fragmentation with night/day activity and streaming bursts. The following helper counts *long inbound streaming* sessions ($\geq$2 h) by first extracting streaming rows and then applying the general sessionisation.
(Listing 5.8, app/metrics/criterion1.py, Appendix C.1, A7)

Listing 5.8: Long inbound streaming sessions; file app/metrics/criterion1.py.

```
5.8.1 #Long inbound streaming sessions ($\ge$2\,h)
5.8.2 def _count_long_streams_ge_2h(df: pd.DataFrame, gap_sec: int = 300)
          -> Tuple[float, Dict[str, Any]]:
5.8.3     rows = df.loc[streaming_inbound_mask(df)] if "Timestamp" in df.
          columns else pd.DataFrame()
5.8.4     if rows.empty:
5.8.5         return 0.0, {"long_streams": 0}
5.8.6     sess = sessions_from_timestamps(rows, gap_sec=gap_sec)
5.8.7     count = sum(1 for (a, b) in sess if (b - a).total_seconds() >=
          7200.0)
5.8.8     return float(count), {"long_streams": int(count)}
```

5.5.4 C2 - Loss of Interest / Anhedonia

One example for C2 emphasizes social engagement versus passive consumption. As a basic building block, I count messaging sessions using the standard gap rule (Listing 5.9).

Listing 5.9: Chat session count; file app/metrics/criterion2.py.

```
5.9.1 def _chat_session_count(df: pd.DataFrame, gap_sec: int = 300) ->
      float:
5.9.2     d = df.loc[chat_mask(df)] if df is not None and not df.empty
      else pd.DataFrame()
5.9.3     if d.empty or "Timestamp" not in d.columns:
5.9.4         return np.nan
5.9.5     sess = sessions_from_timestamps(d, gap_sec=gap_sec)
5.9.6     return float(len(sess)) if sess else 0.0
```

5.5.5 C4 - Sleep Timing & Duration

One metric in C4 derives the onset/wake and fragmentation features from the activity of the header alone. We use a small utility to compute minutes from the 22:00 anchor (Listing 5.10), called after detecting the first $\geq$30 min night gap.

Listing 5.10: Minutes from 22:00 anchor; file app/metrics/criterion4.py.

```
5.10.1 def _minutes_from_2200(ts: pd.Timestamp) -> float:
5.10.2     """Minutes between 22:00 (same local day anchor) and timestamp;
       wraps across midnight."""
5.10.3     anchor = ts.normalize() + pd.Timedelta(hours=22)
5.10.4     if ts < anchor:
5.10.5         ts = ts + pd.Timedelta(days=1)
5.10.6     return float((ts - anchor).total_seconds() / 60.0)
```

5.5.6 C6 - Fatigue / Low Energy

To indicate towards the DSM-5 MDD, Fatigue / low energy C6, we can focus on daytime activation. The following is the computation of the *daytime idle ratio* between 09:00 and –18:00 (Listing 5.11).

Listing 5.11: Day-time idle ratio; file app/metrics/criterion6.py.

```
5.11.1 def _day_idle_ratio(df: pd.DataFrame) -> tuple[float, Dict[str, Any
       ]]:
5.11.2     if df is None or df.empty or "Timestamp" not in df.columns:
5.11.3         return (np.nan, {})
5.11.4     d = df.copy()
5.11.5     d["Timestamp"] = pd.to_datetime(d["Timestamp"], errors="coerce")
5.11.6     d = d.dropna(subset=["Timestamp"]).sort_values("Timestamp")
5.11.7     per_min = d.set_index("Timestamp").assign(cnt=1)["cnt"].resample
       ("1Min").sum()
5.11.8     window = per_min.between_time("09:00", "17:59")
5.11.9     idle = int((window == 0).sum())
5.11.10    ratio = idle / 540.0 if len(window) else np.nan
5.11.11    return (ratio, {"idle_min": idle})
```

5.6 Fuzzy Additive Symptom Likelihood and DSM-5 Gate

This section documents the concrete implementation used in our application to (i) compute daily, criterion-level Fuzzy Additive Symptom Likelihoods (FASL, $L_k \in [0,1]$) from PCAP-derived metrics, and (ii) aggregate those daily likelihoods over a rolling window with a DSM-5–consistent gate rule. The structure aligns with the System Design in Chapter 4.

5.6.1 FASL Scope and Data Flow

For each selected dataset and day, PCAP partitions are loaded, minimally cleaned and enriched with hostnames. From this, a single *daily base record* is computed that downstream criteria reuse. The function `compute_daily_base_record()` in `app/metrics/base_features.py` constructs that record and is the sole producer of the daily `ALL_DAILY` columns consumed by FASL.

To enhance readability, we present below the underlying logic of the FASL and DSM-5 gate as pseudo-algorithms rather than full Python listings, as the original implementation

would be too lengthy and potentially confusing. All the respective Python functions are referenced and can be found in `app/pages/01_FASL_DSM_Gate.py`.

5.6.1.1 Membership Functions and Weighted Aggregation

Each selected metric for a DSM-5 MDD criterion k is mapped to a membership value $\mu_{k,i}(x) \in [0,1]$ using either a triangular or an exponential ramp membership function. These functions are implemented as `mf_tri()` and `mf_exp_ramp()` in `app/pages/01_FASL_DSM_Gate.py`. Metric memberships are then combined additively with their criterion-specific weights $w_{k,i}$ by the function `fasl_score()` to yield the daily likelihood $L_k(t)$.

Algorithm 5.1 Evaluate membership μ for a metric value x (`mf_tri()`, `mf_exp_ramp`)

Require: Metric value x, membership spec $(lo, mid, hi, \text{invert})$, type $\in \{\texttt{tri}, \texttt{exp\_ramp}\}$

5.1.1 **if** x is invalid **then return** 0

5.1.2 **if** type is `tri` **then** compute piecewise linear membership on $[lo, mid, hi]$

5.1.3 **else** compute exponential ramp from lo to hi with midpoint mid

5.1.4 **if** *invert* **then** set $\mu \leftarrow 1 - \mu$

5.1.5 **return** $\text{clip}(\mu, 0, 1)$

Algorithm 5.2 Criterion-level daily FASL likelihood $L_k(t)$ (`fasl_score()`).

Require: Metric values $\{x_i\}$ from the daily record, spec $\{(w_i, \text{MF}_i)\}$

5.2.1 $total \leftarrow 0$

5.2.2 **for** each metric i **do**

5.2.3 $\quad \mu_i \leftarrow \text{EvalMembership}(x_i, \text{MF}_i)$

5.2.4 $\quad total \leftarrow total + w_i \cdot \mu_i$

5.2.5 **return** $\text{clip}(total, 0, 1)$

Algorithm 5.2 corresponds to `fasl_score()` and produces a single daily likelihood $L_k(t)$ for criterion k. The configuration file `fasl_config.json` defines for each criterion the metric set, their weights, and membership parameters. For each day in `ALL_DAILY`, the application builds a values-map that passes to `fasl_score()`, generating the daily likelihood series $L_k(t)$.

5.6.2 Rolling Window and DSM-5 Gate

The DSM-5 gate evaluates whether a criterion is considered present based on its recent daily likelihoods. The function `gate_present()` in `app/pages/01_FASL_DSM_Gate.py` applies the rule: a criterion is *present* if, within the last M days, at least N days have $L_k(t) \geq \theta$. This implements the DSM-5 MDD definition of persistence within an observation window.

Algorithm 5.3 DSM-5 Gate Rule (`gate_present()`)

Require: Sequence $\{L_k(d)\}$, window M, threshold θ, required days N

5.3.1 $S \leftarrow$ last M values of L_k

5.3.2 $c \leftarrow |\{v \in S : v \geq \theta\}|$

5.3.3 **return** $(c \geq N)$

5.6.3 Building and Using the `ALL_DAILY` Table

Algorithm 5.4 outlines how the system creates and maintains the `ALL_DAILY` dataset, which serves as the input for all FASL and DSM-5 MDD computations. It mirrors the implementation found in:
`compute_or_load_all_days_features()` within `app/pages/02_Network_Metrics.py` and `compute_daily_base_record()` in `app/metrics/base_features.py`.

Algorithm 5.4 Constructing `ALL_DAILY` from daily base records

Require: Set of days $\mathscr{D}$, selected datasets

5.4.1 **for** each day $d \in \mathscr{D}$ **do**

5.4.2 Load 5-minute partitions $\rightarrow df_d$

5.4.3 Enrich DNS $\rightarrow$ hostnames, SLDs

5.4.4 $r \leftarrow$ `compute_daily_base_record`(df_d)

5.4.5 **for** each criterion module C (C1–C9) **do**

5.4.6 C.`compute`$(df_d,\ r,\ $`ALL_DAILY`$)$

5.4.7 Append r to `ALL_DAILY`

5.4.8 **return** `ALL_DAILY`

5.6.4 Operational Flow and User Feedback

Once the daily FASL values $L_k(t)$ are available for all criteria, they are used in two ways:

1. The DSM-5 gate (Algorithm 5.3) evaluates the persistence of elevated criteria and determines the presence or absence of each symptom cluster.

2. The same $L_k(t)$ values are visualized on the "Early Signs" dashboard as interactive gauge indicators (Figure 5.2), giving users immediate, criterion-level feedback relative to the threshold θ.

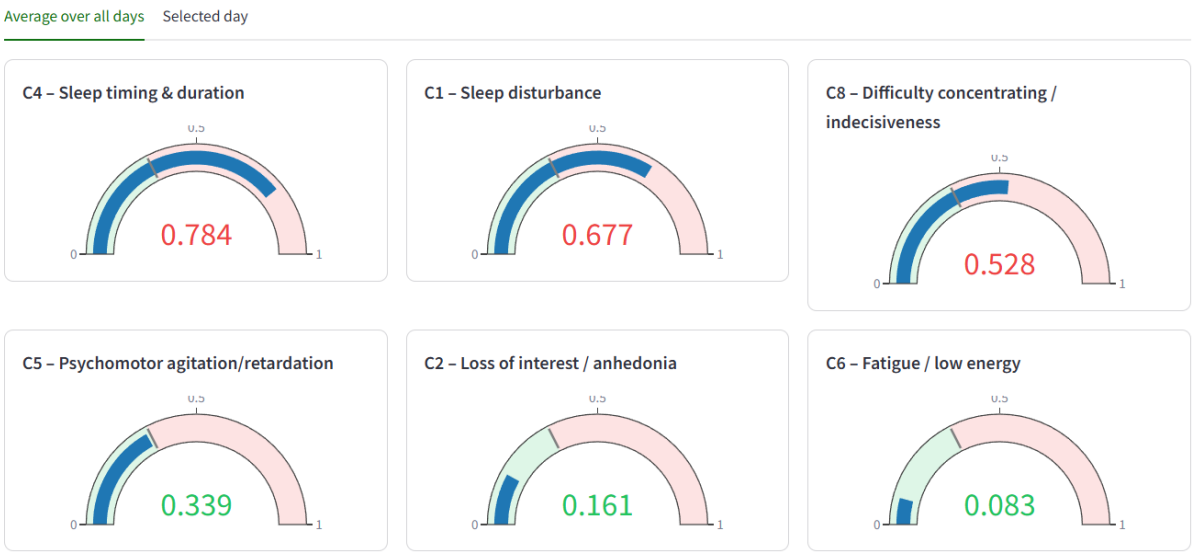


Figure 5.2: Criterion-level FASL indicators showing daily likelihood relative to threshold θ.

5.6.4.1 Metric Configuration

The app ships with an explicit per-criterion dictionary of metrics, weights w, and triangular parameters $(\ell o, mid, hi, invert)$. The configuration is loaded over a JSON file as default direct on the page that computes L_k so the same settings directly drive both computation and visualization. The configuration can be adjusted and downloaded in JSON representation and also be uploaded in a later stage.

Listing 5.12: FASL configuration `fasl_config.json`

```
{
  "M": 14,
  "N": 6,
  "theta": 0.6,
  "core_symptoms": [
    "C2"
  ],
  "C1": {
    "F1_DistinctSocial": {
      "w": 0.25,
      "bom": "",
      "mf": {
        "type": "tri",
        "lo": 20.0,
        "mid": 80.0,
        "hi": 80.0,
        "invert": false
      }
    },
    ...
  ]
}
```

5.6.5 DSM-5 Gate Scope and Data Flow

Before detailing the DSM-5 Gate implementation, it is important to note a deliberate implementation decision. In this prototype, the algorithm for the Behavior Observation Metric (BOM) fusions described in Chapter 4, Section 4.5.3, were not implemented. Instead, a simplified 1:1 mapping was adopted where we selected for each metric that corresponds directly to one DSM-5 MDD criterion, without considering the BOM grouping components. This means that while multiple metrics can support a single criterion, each metric is unique to its observation, and no risk of overweighting arises from overlapping group memberships.

This simplification reduces implementation complexity and keeps the additive FASL

logic transparent while still producing interpretable daily likelihoods for each criterion. The results remain suitable for evaluation in Chapter 6.

5.6.5.1 Gate Function and Episode Decision

After the daily FASL likelihoods $L_k(t)$ are calculated for each DSM-5 MDD criterion, the gate acts as a transparent and rule-based decision layer that evaluates those continuous likelihoods over time. Its purpose is to translate probabilistic daily indications into interpretable binary states, whether a criterion is considered *present* within a clinically motivated window (see more in Chapter 4, Section 4.5.4). The implementation follows this logic closely. A function implemented in `app/pages/01_Early_Signs_Overview.py` iterates over the series of daily likelihoods per-criterion, counts how many exceed θ within the last M days, and marks the criterion as *present* if the count meets or exceeds N. The surrounding code then aggregates these per-criterion results into a single episode-level decision, checking whether at least five criteria are positive and one of them belongs to the configured set of core symptoms.

The function in Algorithm 5.3 takes a pandas series of daily likelihoods for one criterion k and decides if the criterion is *present* at the current time:

1. **Data Hygiene** `dropna().astype(float)` removes missing entries and normalises numerics. This ensures that only valid daily scores enter the gate.
2. **Fixed Horizon** `tail(window)` restricts the evaluation to the last $M =$ `window` days, directly mirroring the DSM-5 "nearly every day" notion.
3. **Empty-Guard** if the windowed series is empty, the function returns `False` (insufficient evidence), so sparse data cannot trigger false positives.
4. **Counting positives** `(s >= theta).sum()` counts how many of the last M days exceeded the threshold θ, the criterion is present if this count is at least $N =$ `need_days`.

In short, Algorithm 5.5 is a literal implementation of the gate rule explained in Chapter 4, Section 4.5.4, with explicit guards for data quality and a deterministic M-day horizon. (`app/pages/01_Early_Signs_Overview.py`, Appendix C.1, A2)

Algorithm 5.6 loops over the nine DSM-5 MDD criteria and builds the boolean dictionary `present`. The interesting implementation details are:

- **Safe column access** that
  `s = DF_L[f"L_{crit}"] if ... else pd.Series(...)` tolerates partially populated data frames (e.g. if a criterion had no configured metrics yet).
- **One source of truth for parameters** so that `theta_default`, `gate_need`, and `gate_window` are read from the page configuration (`cfg_state`), so the same values drive both computation and UI. This keeps runs auditable and reproducible.
- **Core-symptom rule** `core_ok = any(...)` checks whether at least one core criterion (by default including C2) is present.
- **DSM-5 counting rule** `total_present` counts how many criteria are present, the episode flag `episode_likely` is raised only if ≥ 5 criteria are present and a core symptom is present, exactly matching the DSM-5 MDD episode condition.

This separation of concerns keeps the gate logic small and the orchestration transparent. Computationally, each `gate_present` call is $\mathcal{O}(M)$ over the last M values and the full pass across nine criteria are fast enough for interactive updates.

Algorithm 5.5 DSM-5 Gate (`gate_present`)

Require: Sequence $\{L_k(d)\}$ (daily likelihoods), window M, threshold θ, required days N

5.5.1 $S \leftarrow$ last M values of L_k after removing missing entries

5.5.2 $c \leftarrow |\{\, v \in S \mid v \geq \theta \,\}|$

5.5.3 **return** $(c \geq N)$

Algorithm 5.6 MDD Episode Summary (DSM-style gate across C1–C9)

Require: Table DF_L with columns $L_{\texttt{C1}}, \ldots, L_{\texttt{C9}}$ (daily likelihoods); threshold θ; window M; required days N; core-criteria set $\mathscr{C}_{\text{core}}$

5.6.1 $present \leftarrow \{\}$
5.6.2 **for all** $\texttt{crit} \in \{\texttt{C1}, \ldots, \texttt{C9}\}$ **do**
5.6.3 **if** DF_L has column $L_{\texttt{crit}}$ **then**
5.6.4 $S \leftarrow DF_L[L_{\texttt{crit}}]$
5.6.5 **else**
5.6.6 $S \leftarrow$ empty numeric sequence
5.6.7 $present[\texttt{crit}] \leftarrow \texttt{gate\_present}(S, \theta, M, N)$
5.6.8
5.6.9 $core\_ok \leftarrow \exists c \in \mathscr{C}_{\text{core}}$ such that $present[c] = \textbf{true}$
5.6.10 $total\_present \leftarrow |\{ v \in present \mid v = \textbf{true} \}|$
5.6.11 $episode\_likely \leftarrow core\_ok \wedge (total\_present \geq 5)$
5.6.12 **return** *episode_likely*

The output of the DSM-5 Gate is a set of simple, interpretable numbers, whether an episode is likely, whether a core symptom is satisfied, and how many criteria are present within the last M days. These results are immediately visualized on the dashboard as summary tiles for the user (see Figure 5.3).

While this provides a transparent snapshot of the current state, the next logical step is to personalize these thresholds. Different individuals have different lifestyles and baseline behaviors, so the same raw signals may carry different meanings. As already explained in the System Design (Chapter 4), personalization is crucial for avoiding false positives and tailoring the interpretation of network-derived signals.

Figure 5.3: Example Indicator Output of DSM-5 Gate

5.7 User Profiles

To keep the evaluation user-centric even on a shared network, we implemented profile mapping where IP addresses can be associated with a person and typical schedule (regular/home-office weekdays, wake/sleep times). This metadata can improve interpretation of work-hour features and suppresses spurious warnings (right now these metadata are not considered within the metric calculations). Profiles are managed in `app/pages/03_User_and_Network_Settings.py`, Appendix C.1, A4., the dedicated interface shown in Figure 5.15.

Each profile represents an individual network participant and links a set of IP addresses with contextual information such as typical workdays, wake-up and sleep times, and regular activities. These parameters are currently stored in structured JSON form and can be easily adjusted or extended. The goal of this design is to provide richer context for the daily metrics computed by the system, e.g. to better understand why late-night activity occurs or to identify deviations from a user's usual schedule. From an implementation perspective, profiles are stored locally and editable through the Streamlit front-end. Each record includes unique identifiers, personal scheduling metadata, and associated IPs. The system retrieves these profiles at runtime and can use them to annotate or filter metric outputs. Although this logic is not yet active within the current computation pipeline, it provides the foundation for context-aware interpretations that go beyond purely statistical indicators.

External evidence supports this direction, such as Müller et al. who demonstrates that non-intrusive user profiling is feasible even from encrypted smart-home traffic, achieving accuracy in identifying devices and events by leveraging packet timing and sequence patterns in ZigBee networks [35]. Their results strengthen our design choice to add user context on top of router-centric, metadata-only analytics, since our approach generalizes the idea beyond a single protocol to all household traffic while keeping processing local and auditable.

5.8 Dashboard Navigation and User Experience

The follwing section narrates a typical session from the perspective of a non-technical user (e.g., a parent of children/teenager or clinician) who wants a clear picture first and optional technical depth later. The journey begins on the landing page, moves to a high-level overview of early depression indicators and then opens up the metric

workspace. Advanced or technical users can also tune model parameters or upload new, or their own data afterwards.

5.8.1 Login and Introduction

A user logs in once and the session remains cached across all the app pages. The `Home` page (`app/00_Home.py`, Appendix C.1, A1) greets the user with a short disclaimer and a plain-language summary of what the dashboard does and does not do (no content inspection, trends over days are more meaningful than single events). Figure 5.4 shows the login dialog.

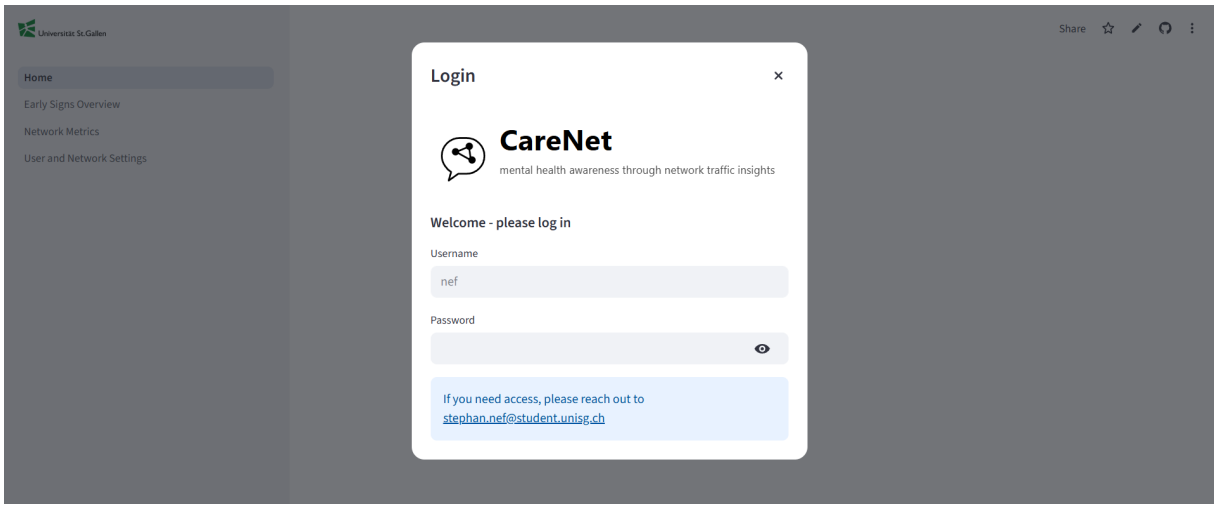


Figure 5.4: Login dialog (session persists for the rest of the app).

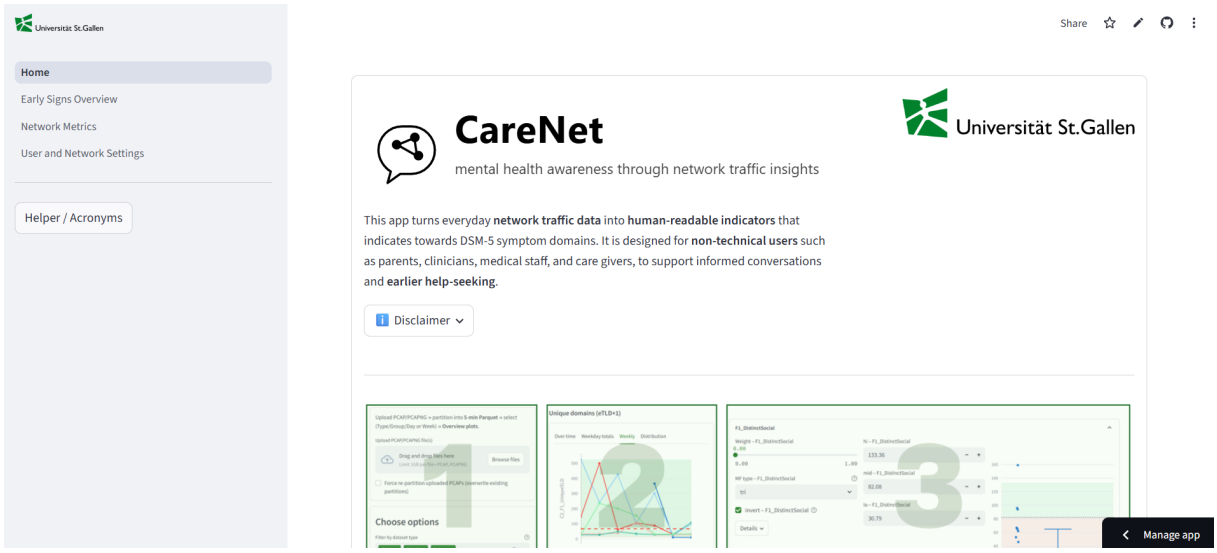


Figure 5.5: Landing page (`Home`) after capturing your screenshot.

5.8.2 Early Signs Overview Page

The `Early Signs Overview` (`app/pages/01_Early_Signs_Overview.py`, Appendix C.1, A2) presents the daily likelihood for each DSM-5 criterion and, above it, the DSM-5 gate status (rolling window of length M that requires at least N days at or above height θ). The view is intentionally simple: green means reassuring, orange requests attention (Figure 5.7). Users can scroll to time-series per-criterion to understand recent trends and how close each curve is to the gate line (Figure 5.8).

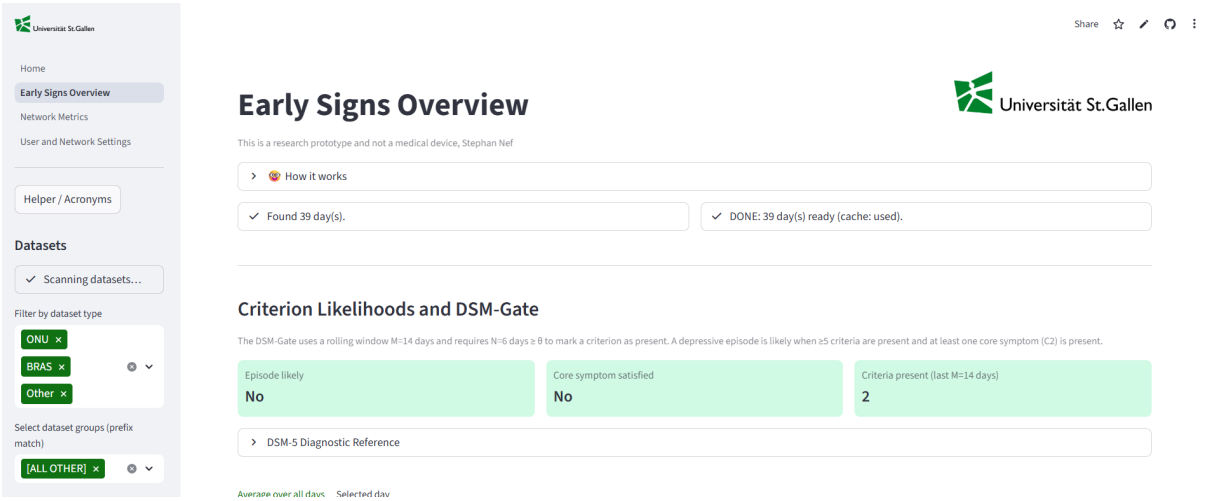


Figure 5.6: Early Signs Overview: current status at a glance.

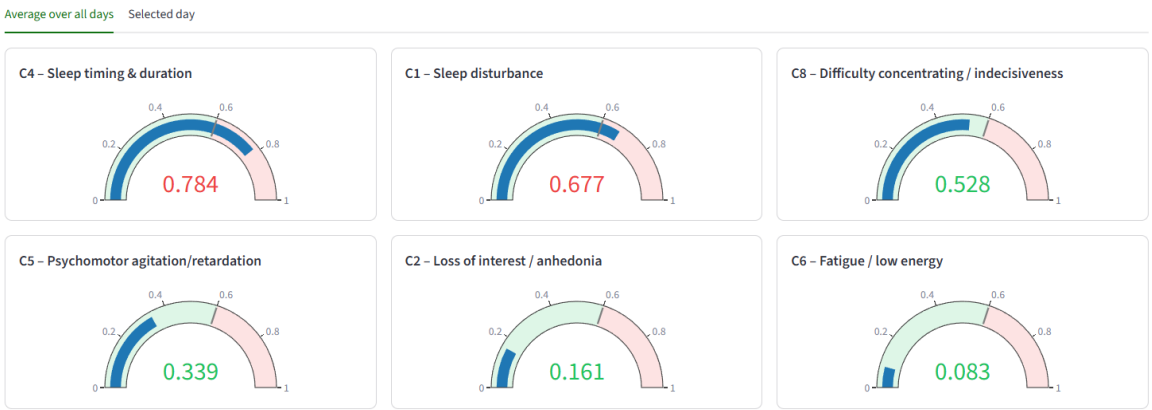


Figure 5.7: Early Signs Overview: per-criterion likelihoods as DSM-5 criterion indicators.

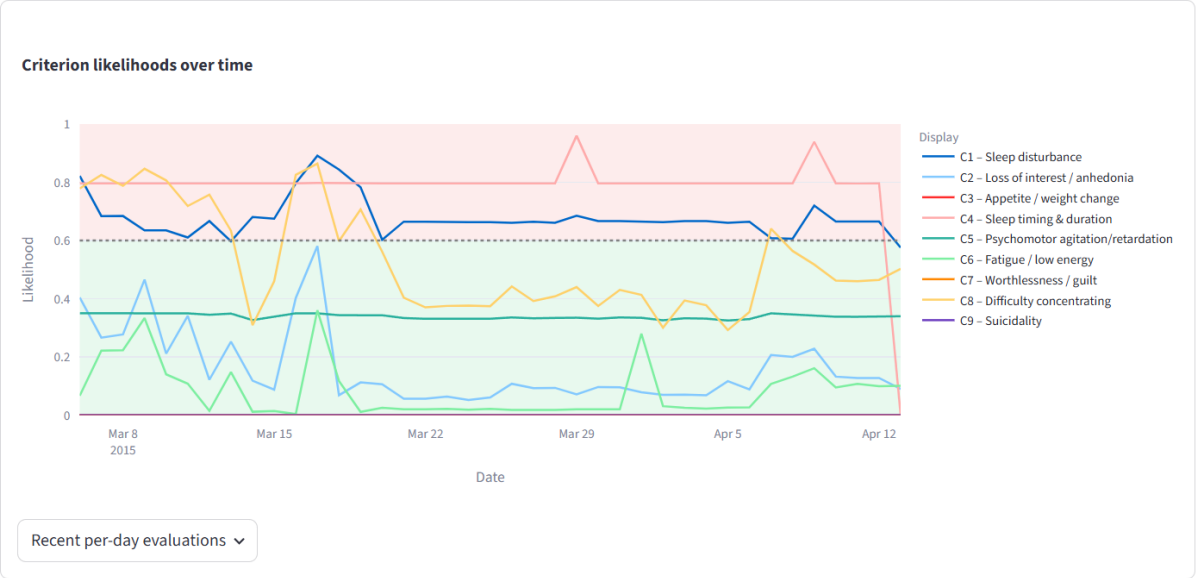


Figure 5.8: Criterion likelihood timelines with the explicit gate threshold line.

All FASL/DSM-5 parameters are visible and editable at the bottom of `Early Signs Overview` page. Non-technical users can leave defaults, while curious users may adjust settings and *export* or *import* them as JSON for reproducibility (Figure 5.7). At the bottom of the page, an advanced panel exposes the exact membership functions (triangular, with an *invert* flag to align direction) and per-metric weights used by FASL. Values can be hand-tuned or suggested from history via a lightweight auto-tune helper, configurations can be downloaded as JSON and re-applied later. Figure 5.9 shows an example parameter card.

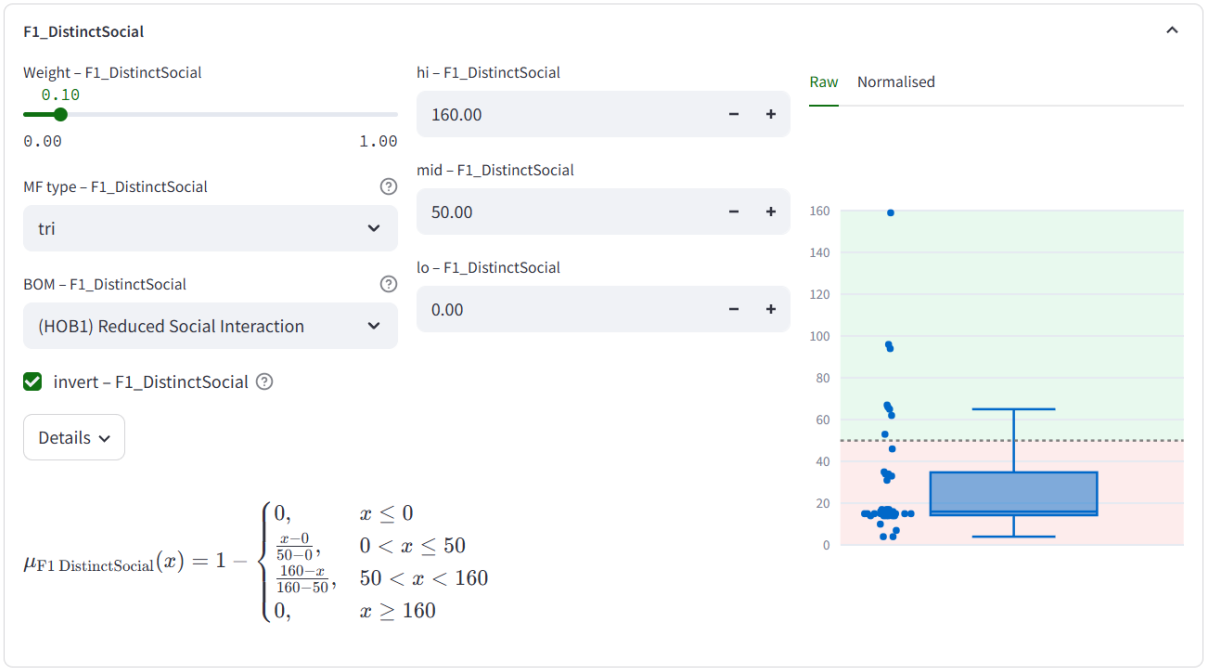


Figure 5.9: Parameter panel: explicit threshold for status, triangular membership, invert function and BOM selection for each metric.

5.8.3 Understand what drives the Indicators

If a user wants to know why a criterion looks elevated, they open `Network Metrics` (`app/pages/02_Network_Metrics.py`, Appendix C.1, A3). For the selected day and dataset, each criterion tab shows a grid of tiles summarising the contributing metrics. Every tile applies the same, transparent rule, one threshold and a direction to derive `OK`/`Caution`. Figure 5.10 illustrates metric grid.

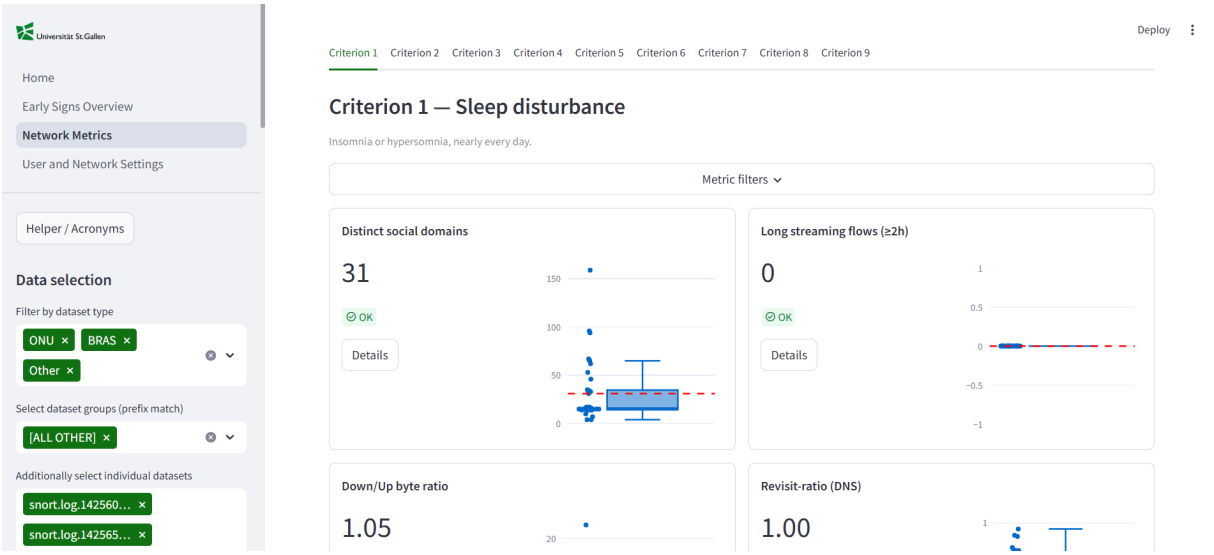


Figure 5.10: Per-criterion metric overview for the selected day.

Clicking the button *Details* turns a single number into context via three complementary views. A distribution boxplot to locate today within the persons own history baseline (Figure 5.11). A weekday/week overlay to expose routine-driven patterns (Figure 5.12) and lastly, an over-time trace to judge persistence and recency (Figure 5.13). The UI also displays the exact LaTeX formula and “numbers in formula” for traceability.

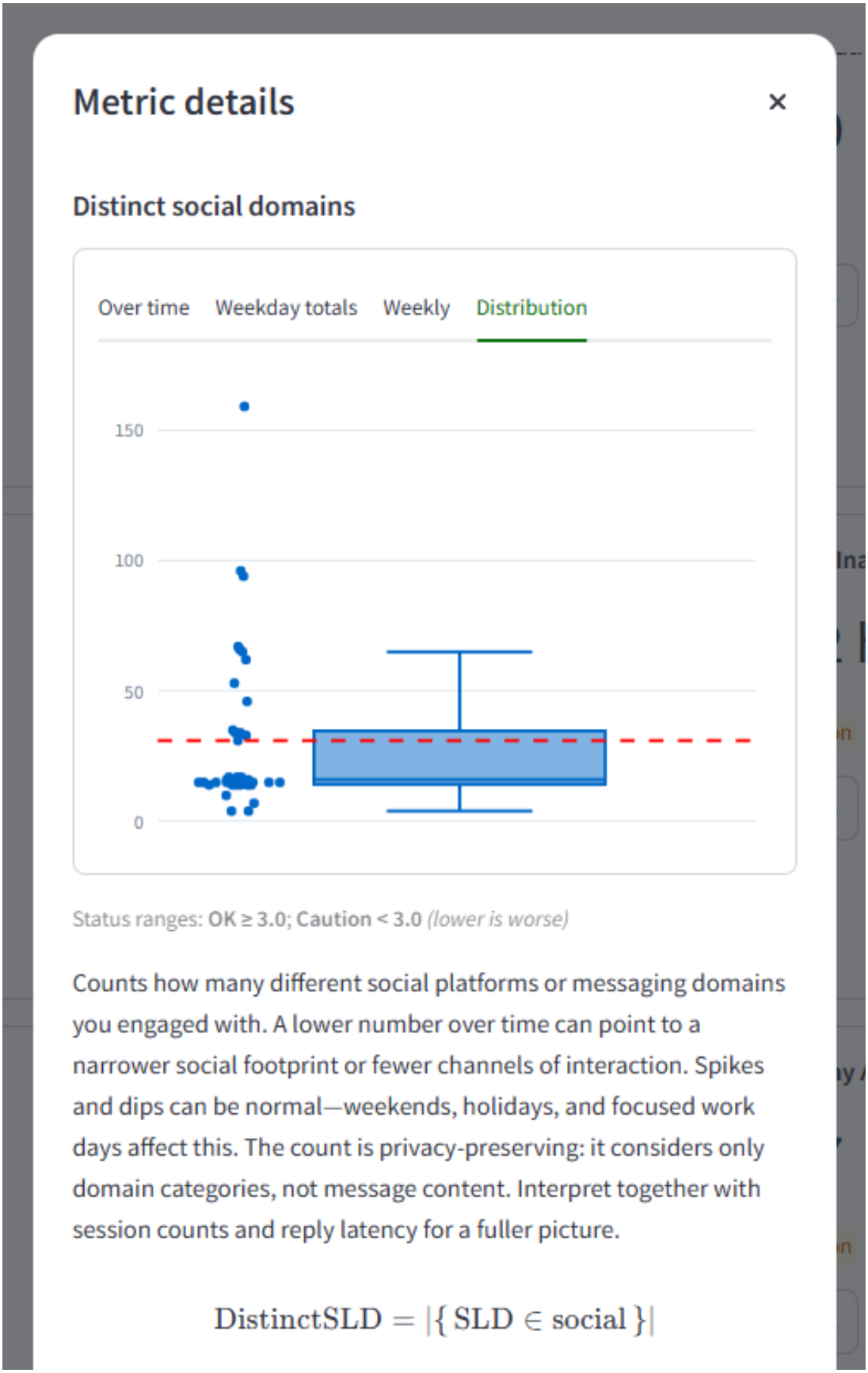


Figure 5.11: Distribution view with the explicit, user-auditable status threshold.

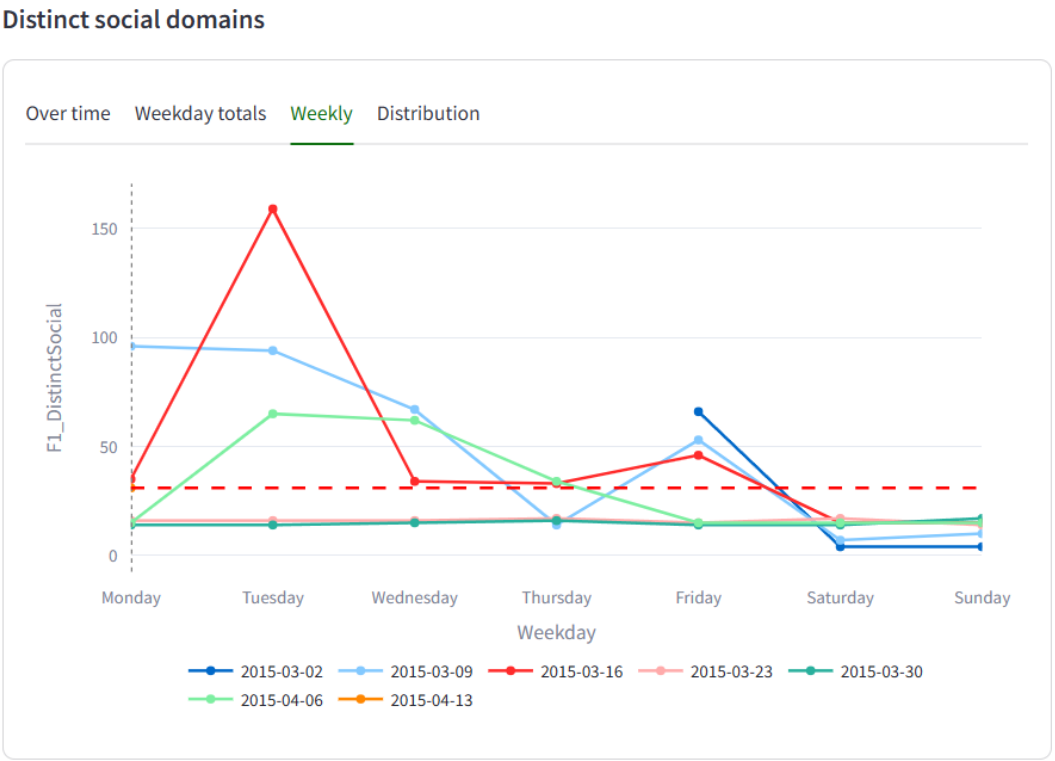


Figure 5.12: Weekday/weekly view to expose routine-dependent differences.

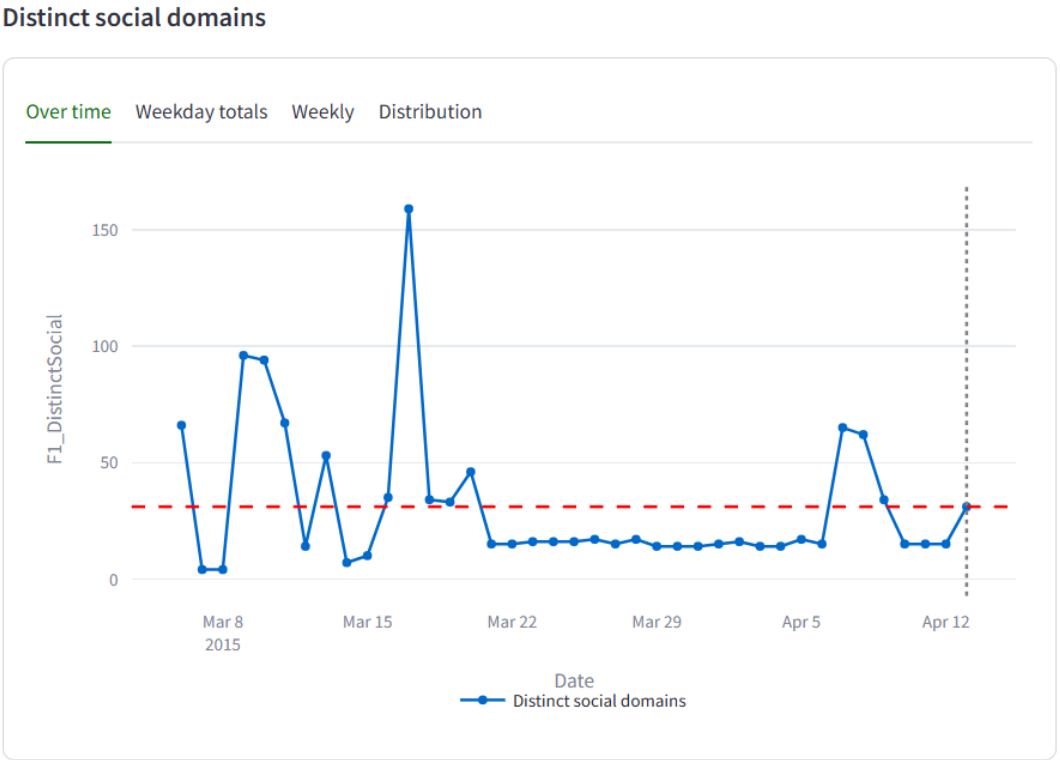


Figure 5.13: Timeline view to assess persistence and recency of changes.

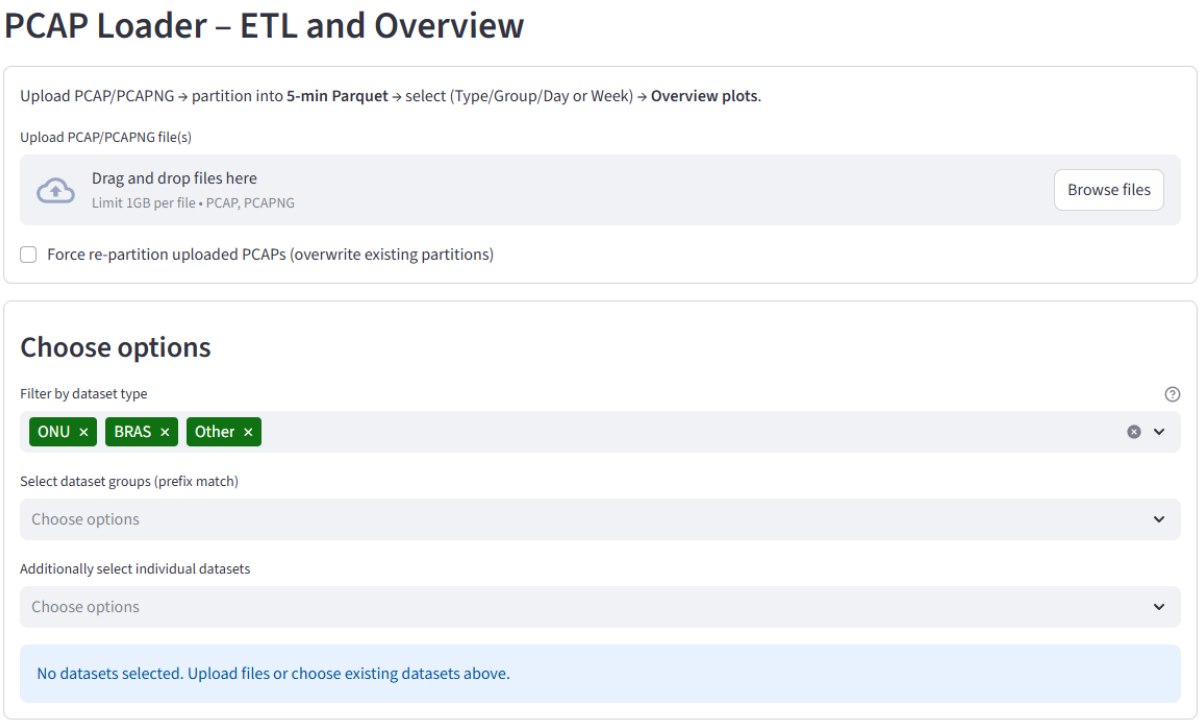


Figure 5.14: PCAP Loader: upload and partitioning to five-minute Parquet with dataset grouping and force re-partition.

5.8.4 Bring your own Data and manage Profiles

After exploring the overview and metrics, technical users can upload their own packet captures on the `PCAP Loader`. One or more PCAP/PCAPNG files are partitioned into contiguous five-minute Parquet slices under `app/processed_parquet/<dataset>`. All datasets are grouped (e.g. ONU/BRAS/Other) and filterable by prefix. A future improvement is a fully automatic capture pipeline that feeds the loader directly (Figure 5.14).

Under `User & Network Settings` you can also maintain profiles (wake/sleep habits, workdays, notes) and map local IPs to users. This mapping activates lightweight direction heuristics (private→public ≈ upstream, public→private ≈ downstream), improving "active vs. passive" features without content access. Profiles have not yet been used within the FASL calculation (Figure 5.15).

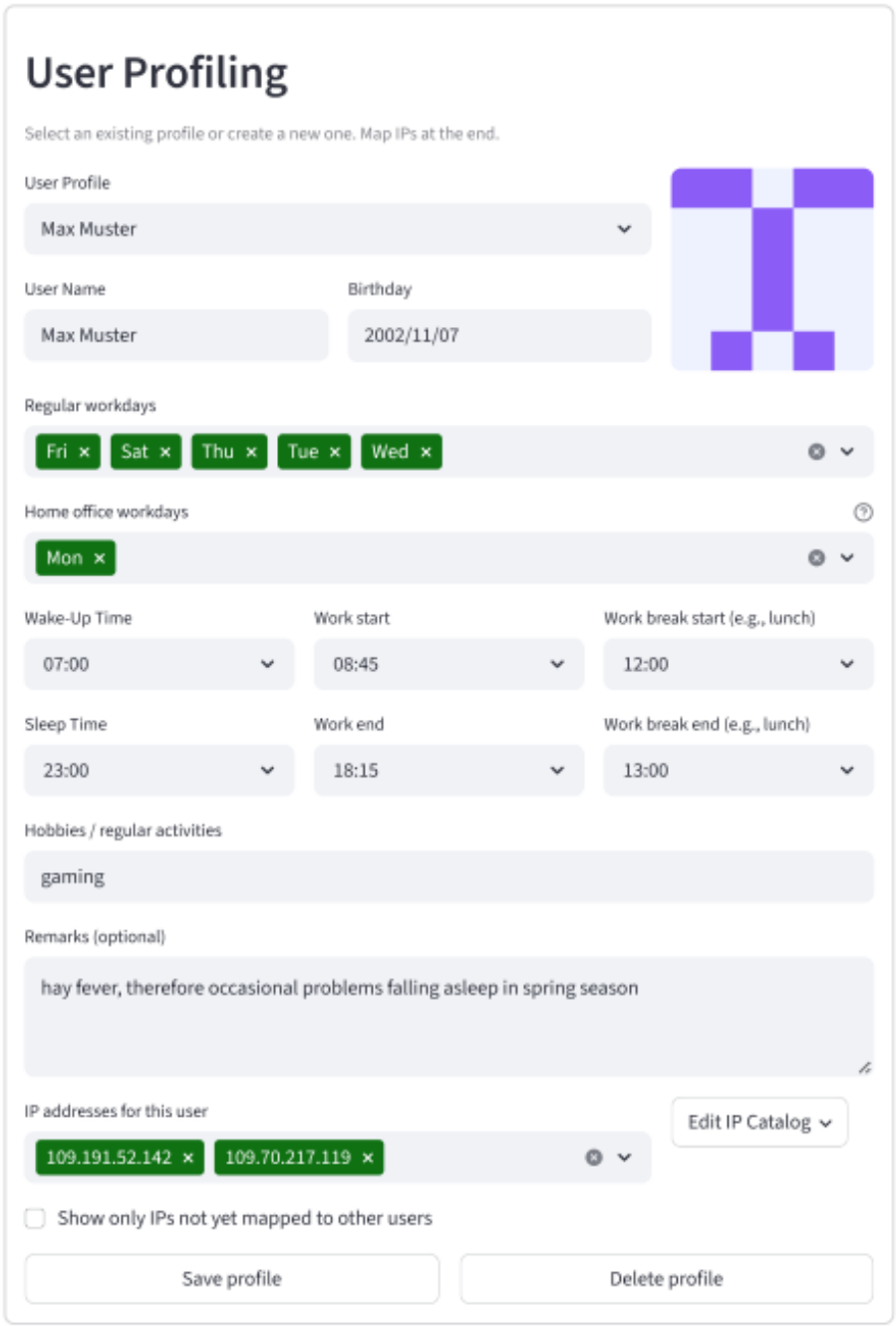


Figure 5.15: User profiling and IP mapping used by direction-aware metrics and interpretations.

Chapter 6

Evaluation

In this Chapter 6 we evaluate whether the implemented prototype and its router-centric data pipeline can support the research question. Specifically, we assess whether the system can reliably extract and aggregate network signals in the Behavior Observation Metric (BOM) in a way that allows the later mapping to indicators of depressive patterns described in the DSM-5 standard described in Chapter 4, Section 4.3.3. To do so, this chapter is structured into two main parts. The first part focuses briefly on high-level technical foundation such as hardware and software setup, deployment environment, and more detailed the datasets that were used to supply the captured network traces. These data and their characteristics are essential to verify that the system design can generate the type and quality of input metrics required for the BOM mapping. The second part, presented in Section 6.2, evaluates some of the calculated metrics and the contributions to the DSM-5 MDD criteria as indicators for depressive episodes.
Together, these two parts provide the empirical basis for judging whether the prototype can answer the research question and providing the groundwork for the following chapters on discussion, conclusion, and future work.

6.1 Evaluation Scenario and Datasets

This first part of the evaluation sets the stage for how the system is judged against the research questions. We describe the hardware and software environment in which the router-centric *Streamlit* application runs, then document the two network traffic datasets that we use throughout the evaluation.

6.1.1 Setup

We established an experimental gateway to validate the system design and to gain a practical understanding of the captured packet data. The setup consisted of a Raspberry Pi 5

with 4 GB RAM connected behind an ASUS RT-AX58U router (Wi-Fi 6, AX3000 class) as a home network environment. We used the Raspberry Pi as a local gateway and began to capture our own packet files (PCAP) for initial inspection. By analyzing these traces with Wireshark, we learned how network packets reflect real interactions. Even without prior technical expertise, Wireshark enabled straightforward inspections such as following TCP streams, examining top talkers, filtering for HTTP requests or inspecting DNS queries. [61]. Soon we recognized that local data collection poses several challenges. The captured files grew rapidly in size even for short recordings, making storage management a major issue. Producing reliable, representative, and unbiased datasets within a home network would require simulating realistic household traffic continuously over weeks, maintaining balanced activity types, labeling each flow correctly and ensuring reproducible storage and rotation. This task would go beyond the scope of this thesis and we decided that this could be pursued as future work. As our research question and system design focuses more on the framework for metric calculation, mapping, and indication rather than on data generation, we decided to rely on publicly available network datasets.

For software development and testing, we built the application on a Lenovo x64 with an AMD Ryzen 7 7730U CPU and containerized it with Docker [16] for portability. The current version was deployed and runs on Streamlit Community Cloud (https://unisg-nef.streamlit.app/), where the computing resources are limited to approximately 0.078–2 vCPUs and 0.69–2.7 GB of RAM with up to 50 GB storage per application [50]. These limits suffice for the extraction, aggregation, and visualization workflows described in this thesis and ensure that all evaluations are reproducible within a controlled environment.

6.1.2 Enabled Datasets

To evaluate the system design and implementation we required network data that are both, labeled e.g. by service type and we needed continuous data over several days. Labeled data allow us to perform content-based analysis such as identifying categories such as *video* or *gaming*, whereas continuous data allow us to compute all time-based metrics such as, e.g., weekday–weekend comparisons and the detection of possible indications of sleep disruption. We did not find any open dataset that fulfilled both criteria, however, we found two complementary ones that we used. Their ingestion and pre-processing steps were already described in Section 5.3.

6.1.2.1 How PCAP data files ties back to the Research Question

PCAP gives packet–level facts that we can transform into the Behavior Observation Metric (BOM) features defined earlier, without relying on opaque aggregates. With PCAP we can derive *n*-minute packet counts, protocol shares, inter-arrival statistics and DNS context which are the building blocks used in our pipeline in Section 5.3 and operationalized in the implemented app described in Chapter 5. We have a Dataset A (6.1.3) that contributes labeled flows that ground content-aware interpretations of BOM features, while another dataset, the Dataset B (6.1.5), contributes continuous, multi-week coverage that supports time-of-day and weekday–weekend analyses. Together they answer the core evaluation need stated in Section 6.1 by linking what happens on the network to BOM indications in a way that is explainable, repeatable, and reproducible.

6.1.3 Dataset A: Real Network Environment Dataset

We use the dataset *A Real Network Environment Dataset for Traffic Analysis* [25]. This dataset couples real home and access traffic with careful processing and service labels from two vantage points (BRAS and ONU). Its labeling yields *>95% application identification accuracy*, giving us reliable ground truth to validate content-aware BOM features and domain/DNS enrichment. In short, Dataset A provides the *semantic backbone* we need to link packet-level descriptors to user activities in an interpretable way. The data were collected in partnership with China Telecom from two vantage points, one at access equipment called Broadband Remote Access Server (BRAS) and one at customer edge equipment called Optical Network Unit (ONU). The BRAS is a central router within the operator's core network that aggregates and manages the internet connections of many households. The ONU, in contrast, is a small device located at the customer premises that converts optical signals into standard network signals for home routers. Collecting data from both sides captures how traffic looks both when it leaves the household and when it enters the wider internet, offering a comprehensive and realistic view of real-world communication patterns.

Figure 6.1 shows the processing steps used by the authors and mirrored in our pipeline. First, traffic is grouped into flows using the five values that identify a connection, namely source and destination addresses, ports, and protocol. Then only the first packets of each flow are kept, and simple side information is computed, such as packet sizes and the time between packets. This step prevents privacy-sensitive content from being released

and compresses the data into stable features that are easy to compare between runs. Service labels are created with a hybrid approach. When traffic is not encrypted, Deep Packet Inspection reads protocol fields such as host names. When traffic is encrypted or otherwise ambiguous, Deep Flow Inspection learns patterns from the side information [25].

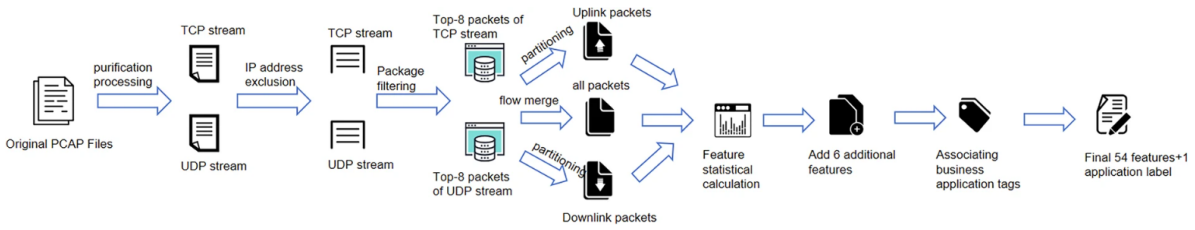


Figure 6.1: Feature extraction process of Real Network Environment Dataset (Figure from [25, Fig. 3]). The process splits PCAP flows by five-tuple, retains the first packets per flow, and computes side-channel statistics used for labeling.

This dataset is a good match for our thesis because it bridges controlled experimental design with the variability of real-world traffic. It includes many users, diverse applications and different service types under real operating conditions. Also, yet remains structured enough to support reproducible analysis. For now, we use this collected dataset as the basis for our evaluation, while the described data processing, especially the Deep Flow Inspection that learns patterns from the side information approach, could be adapted in future work to extend our own privacy preserved data collection and feature extraction pipeline.

As shown in Figure 6.2, the resulting dataset contains labeled traffic sessions for application categories such as *web-browsing*, *video*, *game*, *instant-message* and others. The dataset includes raw `PCAP` files and processed CSV feature tables for both BRAS and ONU sources. Figure 6.2 shows the overall traffic distribution, where video and network-transmission dominate in packet and byte count, followed by gaming and web browsing. This heterogeneity supports content-based interpretation of Behavior Observation Metric (BOM) results.

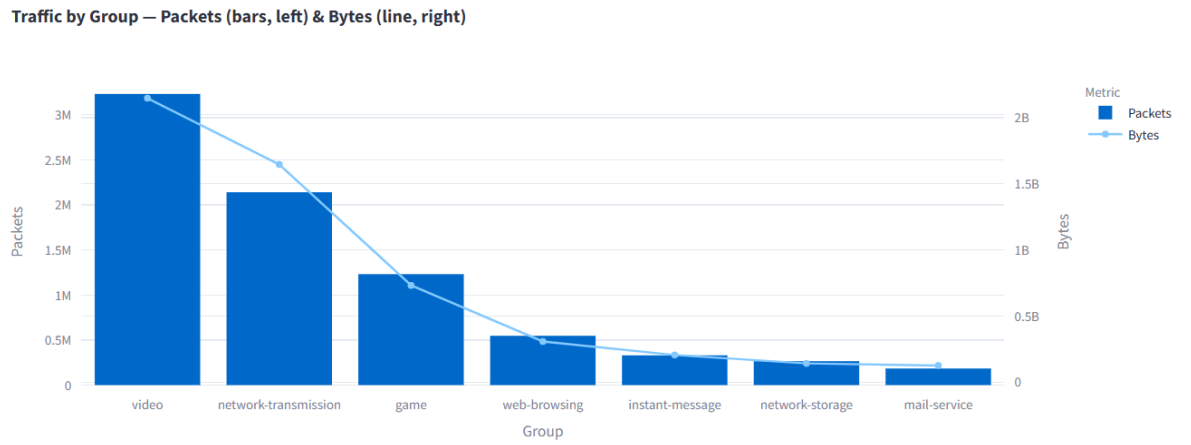


Figure 6.2: Dataset A (Real Network Environment Dataset): Traffic volume across BRAS and ONU by all labeled groups.

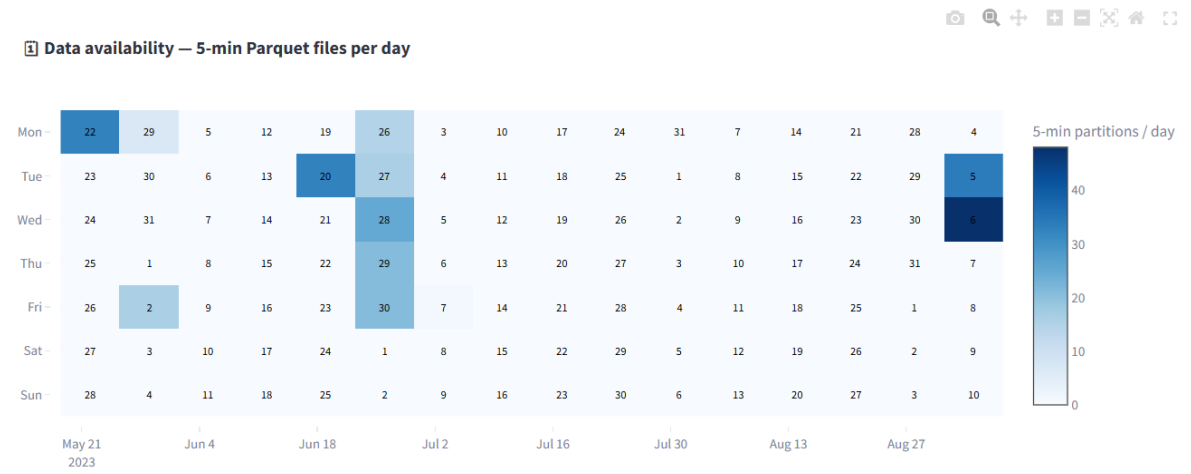


Figure 6.3: Dataset A (Real Network Environment Dataset): Availability across observation days. Gaps indicate that captures were performed on selected dates only, limiting long-term analyses that would be needed to indicate a depressive episode according to the DSM-5 standard.

6.1.4 Protocol and DNS characteristics by Category

We examine the protocol mix at one minute resolution and we relate these patterns to what the DNS enriched data show for the same days. We use the transport layer coding defined by the dataset creators where TCP is coded as 0, UDP as 1, and any other protocol family as 2, which appears as *Other* in our plots [25]. This section focuses on an excerpt of four representative groups that are also available as selectable dataset groups in our

app, namely web browsing, instant messaging, video, and games (see captured services in Table 6.1 from [25, Tab. 3].

Application Category	Application Instance
network-storage	Baidu Netdisk Tianyi Cloud Disk Alibaba Cloud Hua Weiyun
network-transmission	Thunderbolt BT Download Emule
video	Station B Tiktok Tencent Video IQiyi Video Youku Video Mango TV Tencent Meeting
game	League of Legends Honor of Kings
instant-message	QQ WeChat
web-browsing	Alipay webpage Today's Headlines Zhihu Baidu Baike
mail-service	163 email QQ email 189 email

Table 6.1: Service classification labels and instances (reproduced from Table 3 in [25]).

6.1.4.1 Web browsing

Figure 6.4 shows the minute level packet counts over the day, for two adjacent dates,05.-06.09.2023. We see a quiet baseline during the night and early morning, short bursts around 05:00, and a narrow spike around 10:06. Such spikes could indicate for short lived downloads or update checks, which aligns with the expectation that web browsing often happens in short sessions rather than in long continuous flows. For BOM this matters because the features that measure burstiness, peak to median ratios, and time of day distributions are designed to distinguish brief activity windows from extended sessions.

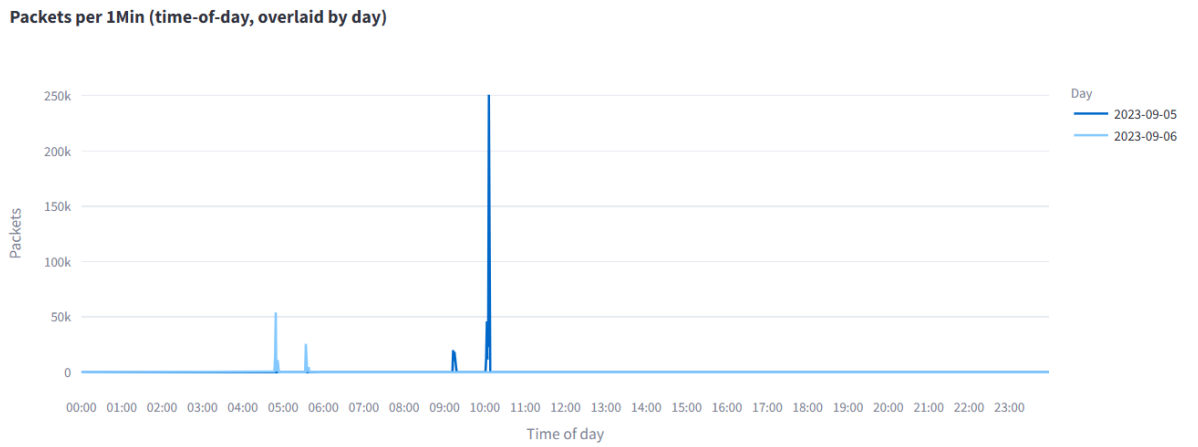


Figure 6.4: Web browsing packets per minute over the day for two adjacent dates. Narrow spikes indicate short-lived activity windows (could indicate downloads or update checks).

The protocol mix for web browsing these days is dominated by TCP, with small contributions in the *Other* category and only marginal UDP, as shown in Figure 6.5. This is consistent with encrypted HTTPS transfers [25]. In the DNS enriched parsed data for the same windows we observe many short name lookups that co-occur with the TCP peaks. This pattern is expected when a browser resolves multiple hostnames for page assets and content delivery networks. For our feature set this is useful because the fraction of TCP packets, the count of distinct DNS names per five minute partition, and the alignment of name lookups with packet spikes help us separate interactive browsing from background noise.

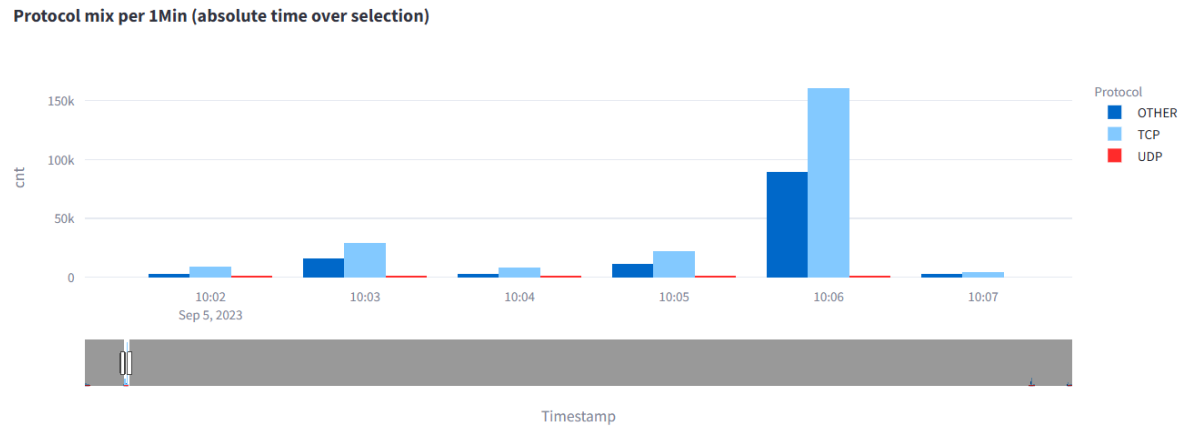


Figure 6.5: Dataset A (Real Network Environment Dataset): Web browsing protocol mix per minute for 05.09.2025, zoom into time 10:02–10:07. *TCP* dominates, *UDP* is marginal, *Other* is present only around bursts.

6.1.4.2 Instant Messaging

Figure 6.6 shows around 5 minutes instant-messaging traffic (06.09.2023). We see short bursts where UDP dominates and TCP stays low-typical of real-time actions such as voice or video calls rather than plain text chat. In the DNS-enriched data around these bursts we also find the expected lookups for instant-messaging services, `QQ` and `WeChat`, which this dataset lists under the instant-message category. In our BOM features this appears as a higher share of UDP packets together with short inter-arrival times, helping us distinguish brief social communication spikes from passive use.

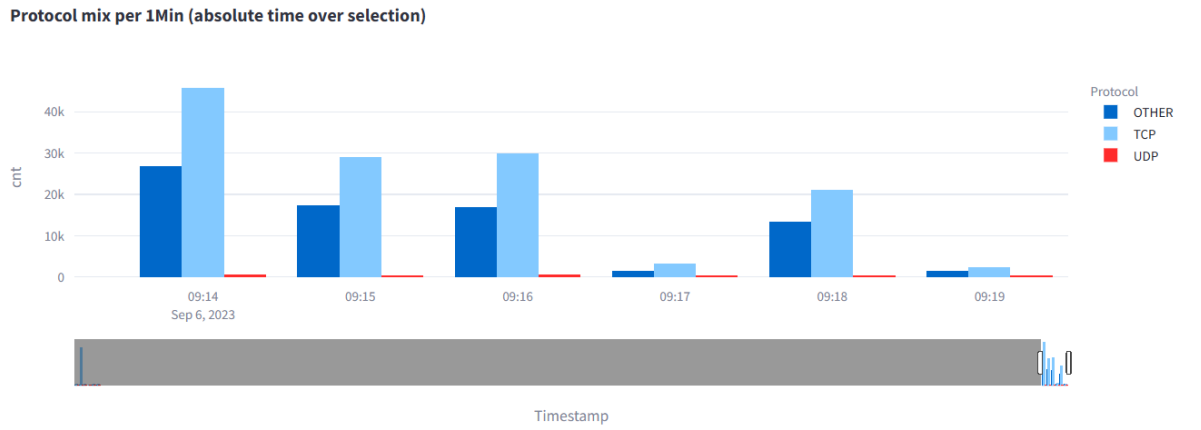


Figure 6.6: Dataset A (Real Network Environment Dataset): Instant messaging protocol mix per minute for 06.09.2023, zoom into time 09:14–09:19. UDP-dominated bursts could also indicate real-time interaction periods. A smaller increment than one minute would be needed to better recognize instant messaging.

6.1.4.3 Video

Figure 6.7 plots the protocol mix for the video group on 05.09.2023. We observe large TCP bars and a smaller contribution in *Other*. This is compatible with encrypted HTTP-based streaming and with the dataset's transport coding that groups non TCP or UDP traffic into *Other* [25]. The DNS enriched data for the same day shows many repeated name resolutions within short time windows, a footprint that often comes from adaptive streaming clients that fetch segments and check multiple content hosts. For BOM this is relevant because the features derived from the packet length distribution and the regularity of segment like bursts are intended to capture sustained media consumption windows. These windows contribute to time of day profiles that we later compare against sleep and rest periods.

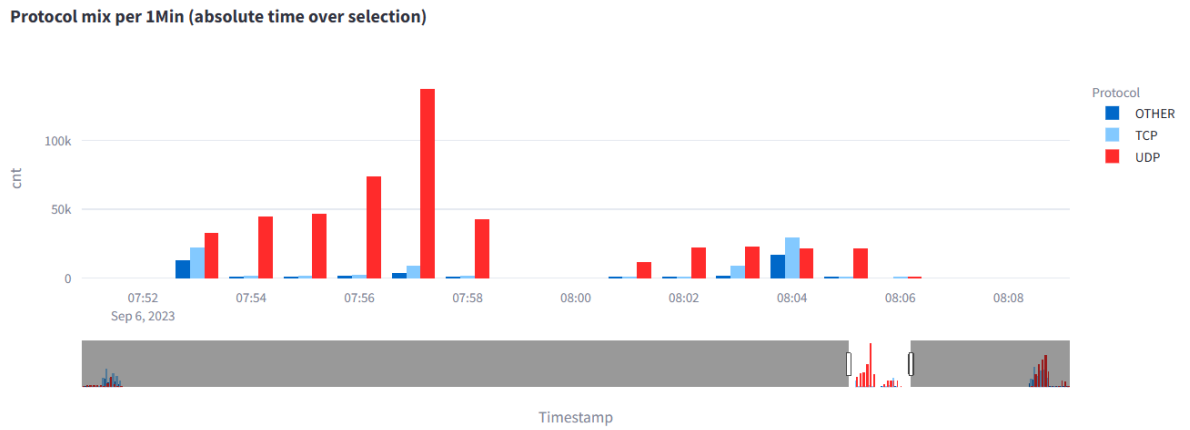


Figure 6.7: Dataset A (Real Network Environment Dataset): Video protocol mix per minute for 05.09.2023, zoom-in to time 07:52–08:08. TCP dominates with a smaller *Other* share, which matches encrypted streaming over HTTPS.

While this dataset provides rich semantic labels, it does not offer continuous coverage across multiple full weeks, as shown in Figure 6.3. The captures were recorded in distinct time windows during the morning, afternoon, and evening hours. Consequently, analyses that rely on day–night continuity such as sleep-disturbance detection cannot be performed reliably. Nevertheless, this dataset serves as a valuable reference for interpreting content-based variations within the BOM framework.

6.1.5 Dataset B: Hands-on Network Forensics Dataset

The second dataset we used comes from the *Hands-on Network Forensics* workshop organized by Erik Hjelmvik from the Swedish Armed Forces CERT back in 2015 [20]. This capture offers a *continuous 40-day* packet stream recorded from a realistic, multi-application environment. Because it preserves raw PCAP files, all time-based features can be recomputed from first principles. Dataset B therefore supplies the *temporal backbone*: robust day–night contrasts, weekday–weekend rhythms, and burstiness-precisely the ingredients needed for our DSM-5 MDD style aggregation and rolling-window gate.
Traffic was captured with the Security Onion platform, which continuously monitors network interfaces and records the raw packet stream into files named `snort.log`. Although these files originate from Snort [48], they contain complete packet captures in standard PCAP format rather than alert messages. Security Onion also integrates Zeek (formerly Bro) [63] and Argus [9] to produce summarized flow and protocol logs, but in our work we directly use `snort.log` PCAP data as the primary data source. This ensures that our

analysis starts from the original packets and remains independent of any pre-computed labels or flow aggregation.

The `snort.log` files are the raw footage from which all later analysis can be reproduced. The capture spans a continuous 40-day period, allowing us to examine long-term behavior trends such as daily or weekly usage cycles. Background traffic includes web browsing (Facebook, search engines) chat applications (Skype, HipChat), email, and cloud services (Dropbox), providing the natural variability needed to test time-based metrics of the BOM.

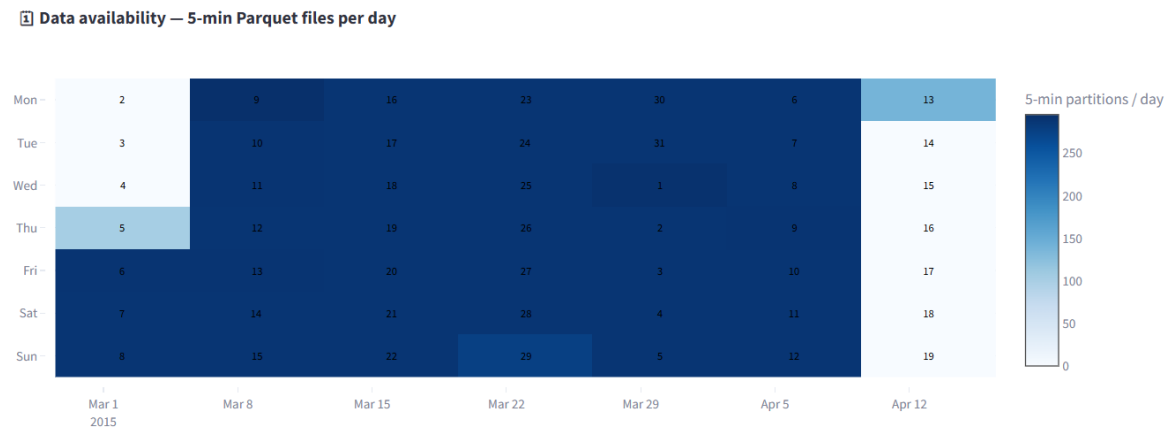


Figure 6.8: Dataset B (Hands-on Network Forensics): Overall data availability showing continuous capture over 40 days.

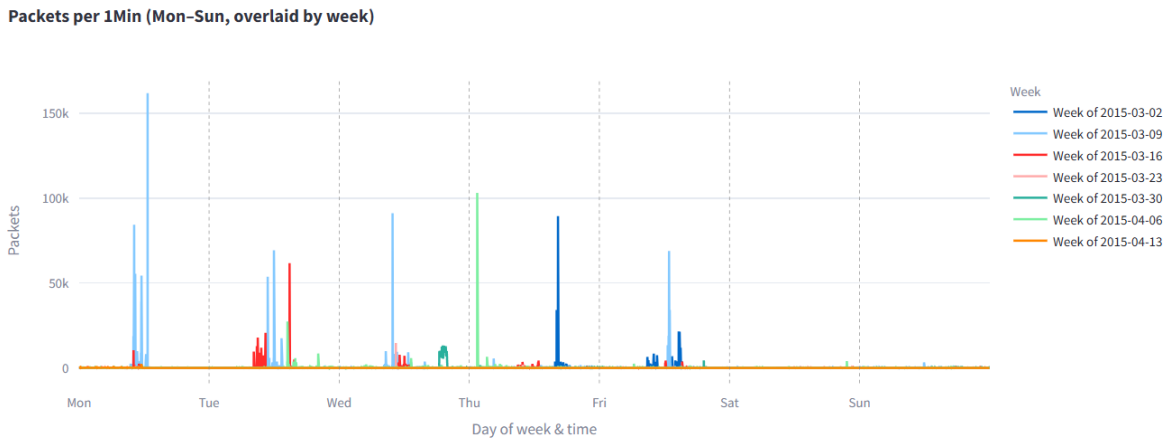


Figure 6.9: Dataset B (Hands-on Network Forensics): Packets per minute by weekday. Regular daily cycles are visible, which support temporal analysis. Some peaks appear Monday to Friday, with none on the weekend (Saturday–Sunday).

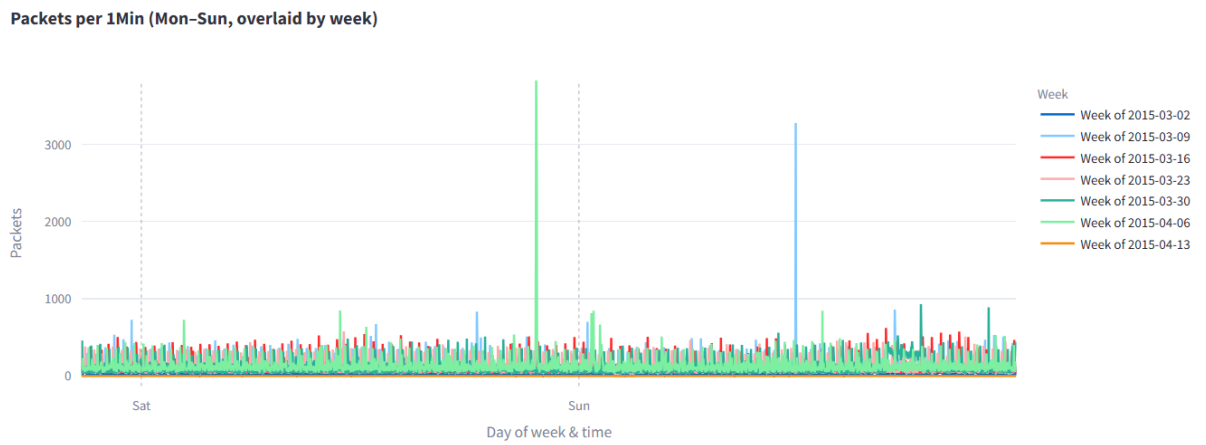


Figure 6.10: Dataset B (Hands-on Network Forensics): Weekend zoom (Saturday–Sunday) showing fluctuating but low activity, consistently below 500 packets per minute.

Figures 6.8, 6.9 and 6.10 show that even on close inspection packet counts fluctuate between 0 and 500 per minute with almost no minutes above 500. The patterns appear random as expected and no individual user behavior can be read from it. However, this makes the dataset still suitable for computing time-based metrics (rather than inferring routines) and we expect clear indications to come from day–night contrasts such as quiet hours that correspond to rest or sleep windows. The capture spans over consistent 40 days which supports stable, random baseline estimation and coverage of weekly cycles.

Before diving into the protocol composition, Figure 6.11 demonstrates that the network capture for 12.04.2025 contains valid data across all hours of the day. Unlike Dataset A, which was limited to a few selected time windows, Dataset B provides a continuous time-series that covers full days and nights without interruptions. This allows us to compute diurnal metrics-day–night contrasts, weekday–weekend differences, and quiet-hour ratios-as proxies for rest or inactivity. However, because we do not observe pronounced overnight lulls, the trace does not resemble typical home-router traffic where user routines are visible in the packet counts.

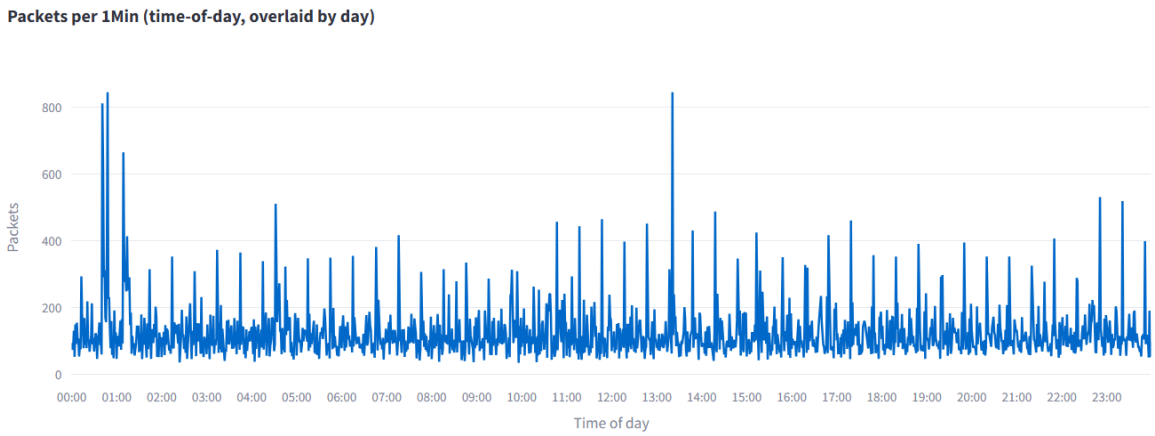


Figure 6.11: Dataset B (Hands-on Network Forensics): Packets per minute on 12.04.2025 shows continuous packet activity throughout the entire 24-hour period without major recording gaps.

6.1.6 Protocol mix and DNS-enriched Data

To complement the high-level description of Dataset B, we analyze one representative day from the continuous capture window, 12.04.2025, and relate packet-level observations to the parsed data available in the app as *Raw Data Preview* and *DNS Enrichment Map*. The aim in this section is to understand what the data contain and how the protocol mix looks over the day.

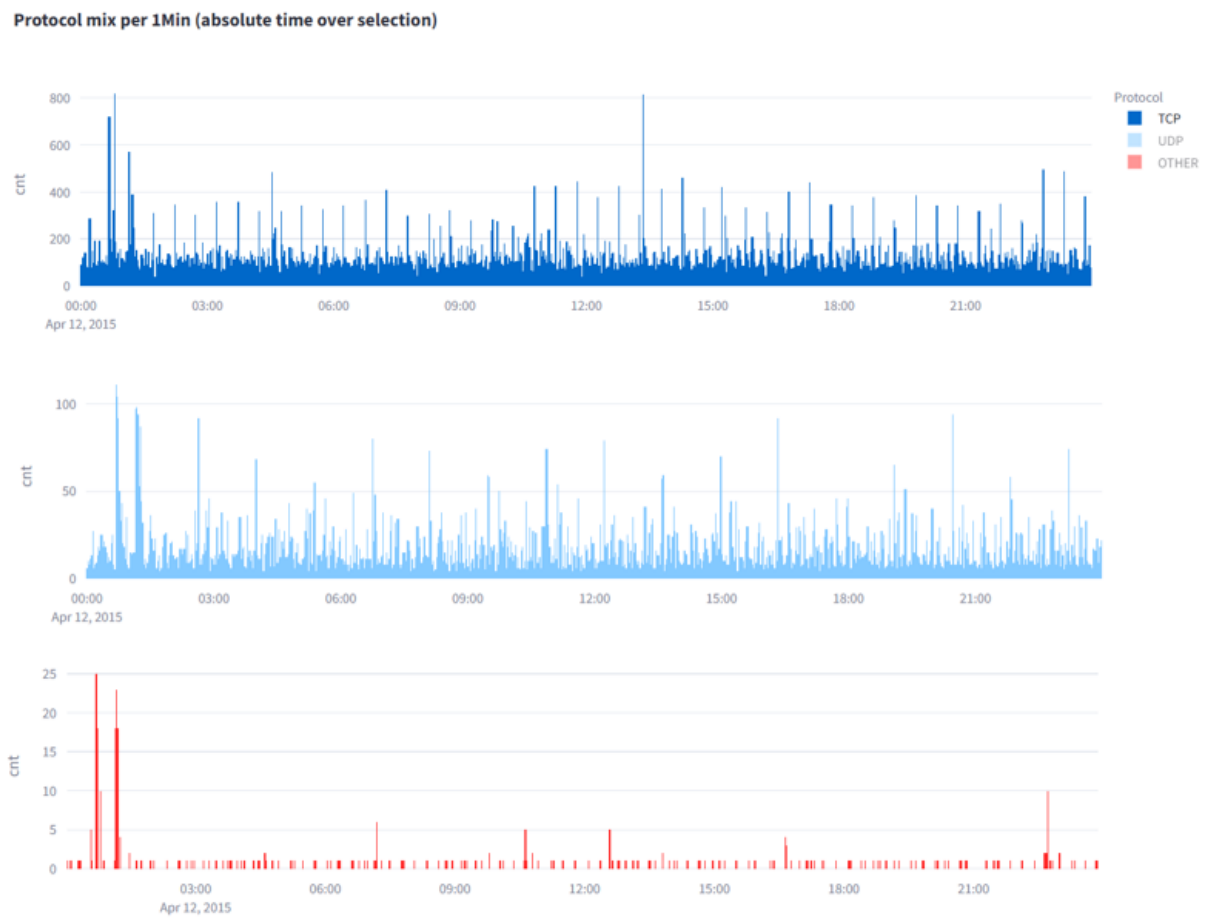


Figure 6.12: Dataset B (Hands-on Network Forensics): Protocol mix per minute on 12.04.2025. The top band shows TCP, the middle band shows UDP, and the bottom band shows traffic classified as Other. The pattern is dominated by low-to-moderate TCP activity with small UDP bursts and only a few minutes with Other traffic.

Figure 6.12 visualizes again the per-minute counts split by transport protocol. TCP forms a steady baseline with frequent but small spikes that appear regularly during the day. UDP appears as short bursts superposed on the baseline, and the Other category is almost absent. This shape is consistent with a small office or home network where web browsing, updates, and email form the majority of traffic, while short UDP bursts often correspond to DNS lookup or brief real-time interactions. This reading aligns with the workshop description of background activity that includes web browsing, chat, email, and Dropbox, interleaved with staged incidents that are sparse in time [20]. For BOM this means that day–night contrasts, burstiness and peak-to-median features are expected to be informative, while content-based interpretation is inherently limited because Dataset B is unlabeled at the application level.

Even without service labels per flow, Dataset B helps to answer the time-based parts of our research questions. The continuous 40-day capture enables stable estimates of day–night activity, weekday–weekend differences, and peak-to-median ratios. The protocol plot in Figure 6.12 shows a low overnight baseline and daytime bursts, which feeds directly into BOM features such as quiet-hour detection and burstiness. The *DNS Enrichment Map* provides counts and first-to-last-seen windows for second-level domains, which we use as a simple, privacy-preserving proxy for breadth of online activity. All dataset explorations and analyses described here are reproducible within the implemented and deployed application (https://unisg-nef.streamlit.app/) [36]. The `User_and_Network_Settings` page allows multi-selecting of `BRAS`, `ONU` and `snort.log` as well as individual available dataset groups such as *video*, *web-browsing*, *gaming*, and more. Users can select specific days or full weeks to reproduce the same visualizations. The interface also provides a *Raw Data Preview* and the option to download the corresponding CSV file for inspection. Figure 6.13 shows this configuration interface.

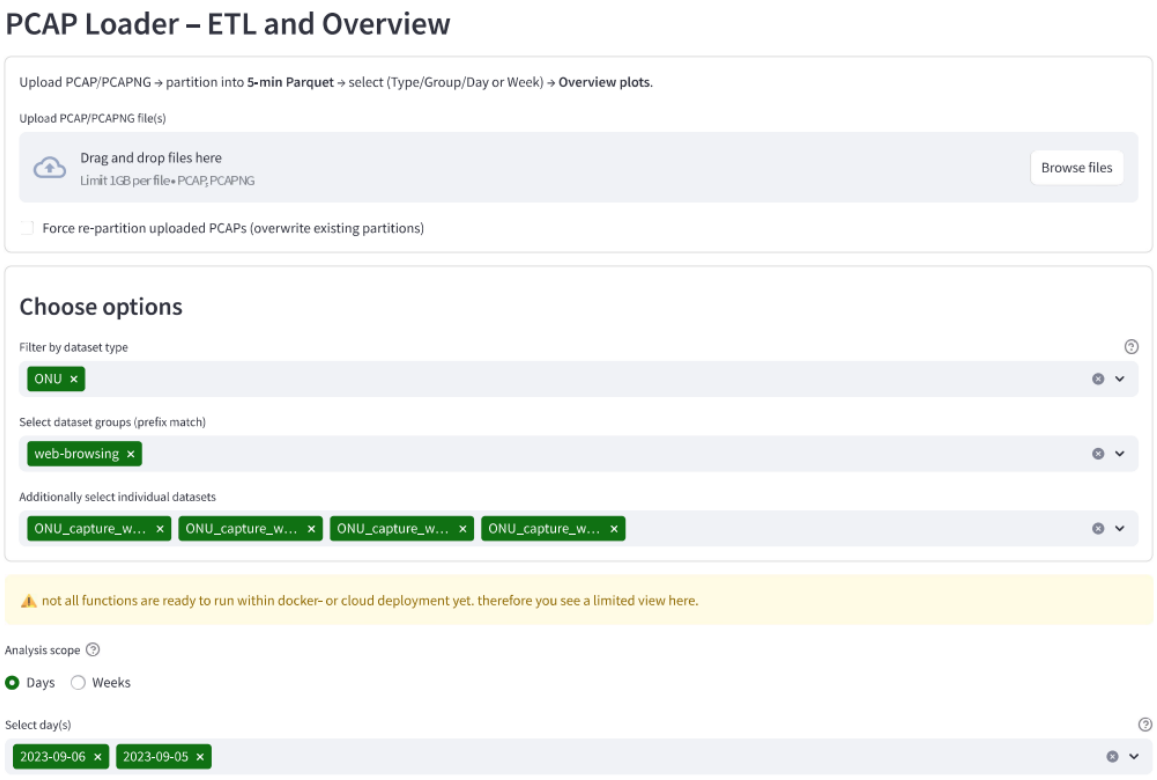


Figure 6.13: App configuration in page `User_and_Network_Settings` for selecting datasets and analysis periods.

6.2 From Packet-Level Metrics to DSM-5 MDD Indicators

In this second part of the evaluation, we proceed bottom up. Means we start from the packet-level metrics that the router-centric system design allows to compute, and we also evaluate how these metrics are directly quantified towards selected DSM-5 MDD criteria (direct because we bypassed the mapping via the BOM in the implementation). The goal in this evaluation is not to cover every DSM-5 criterion, but to check how our current implementation supports with meaningful and reproducible metrics and membership functions. Several DSM-5 MDD symptoms would require signals that a home-router vantage does not provide or that the metrics proposed in the system design chapter do not yet capture robustly. We therefore focus on a small set where the evidence is strongest in our data and code and we treat the other criteria as out of scope for this iteration.

With the two datasets enabled in Section 6.1 and the common ETL in Section 5.3, we materialize five-minute partitions for packet counts, protocol shares, session gaps and DNS context. We use data set A when label view - helps to interpret a metric, and Dataset B when continuous coverage is needed for the diurnal and weekly patterns. Formal parameter definitions used within this chapter, like weights, thresholds, and membership functions are given in Chapter 4. For the following metrics we only use metrics from DSM-5 MDD Criterion 1 (B.1), Criterion 2 (B.2), Criterion 4 (B.4) and Criterion 8 (B.8) to keep it straight forward and focus on the essentials. The 40 day, day-by-day Dataset B from the Swedish Hands-on Network Forensics workshop described in Section 6.1.5 provides a continuous timeline that is well suited for this style of evaluation, although it does not reflect the behavior of one real person because services generate traffic around the clock.

6.2.1 C4_F8 Night/Day Ratio (bytes)

Figure 6.14 shows the raw byte ratio by date. The green band acts as the healthy baseline, and the red band marks the DSM-5 MDD indicator region. These parameters are defined for each metrics as shown in the Figure 6.16 Most days sit between 0.0 and 0.6, which is visually stable over weeks. A single large spike around April 9 reaches beyond value 4. Because that value sits far above the `hi` limit, it will be intentionally ignored by choosing a membership curve and limits that set the indicator band to end below that spike. This demonstrates a practical advantage of the design: the user of the system can remove one-off artifacts without writing rules, simply by positioning the limits.

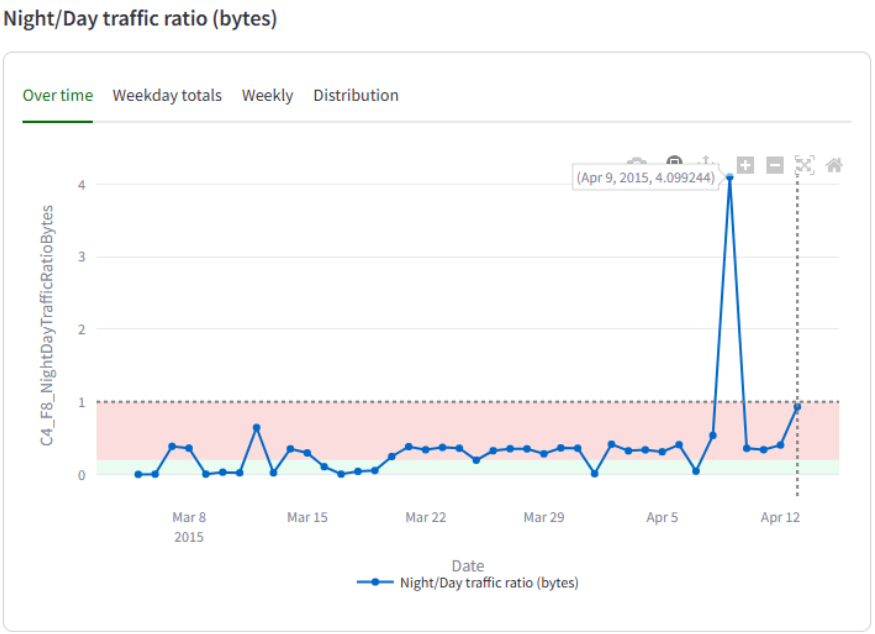


Figure 6.14: C4_F8 Night/Day traffic ratio (bytes), ratio values over time. We can detect one single outlier around April 9.

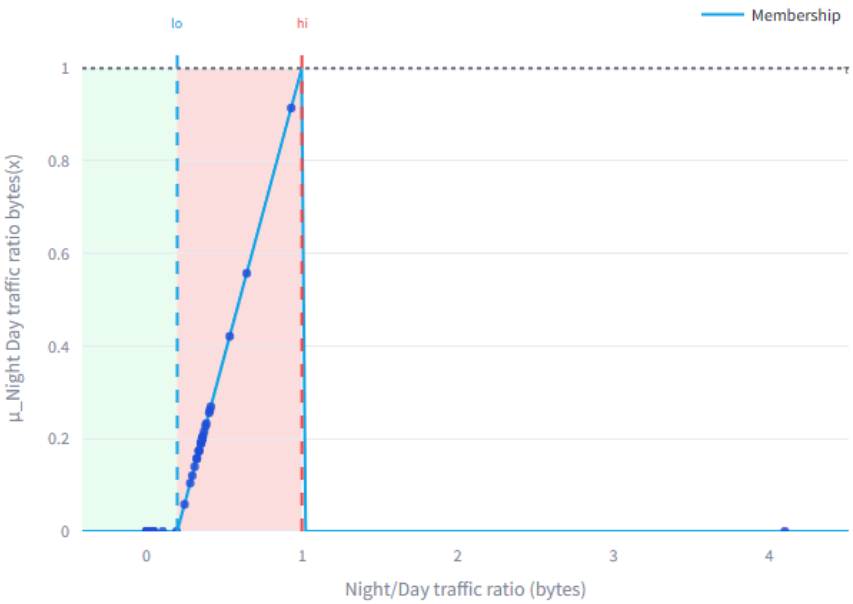


Figure 6.15: C4_F8 Night/Day traffic ratio (bytes), membership function. Triangular function output with a 'right shoulder' (`mid=hi`) and an ignored outlier around 4.0.

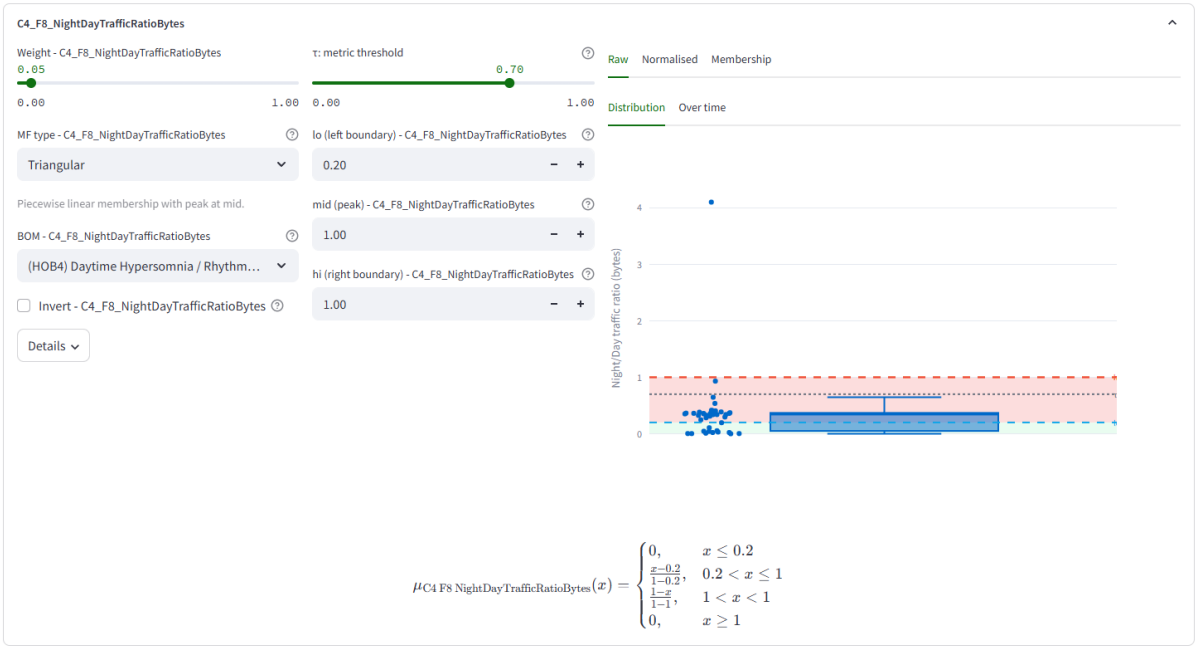


Figure 6.16: C4_F8 Night/Day traffic ratio (bytes), individual parameter settings for each metric. The baseline cluster is stable, while a single outlier around April 9 exceeds the band and therefore will be excluded by the set `hi` limit.

Figures 6.19 and 6.20 (C1_F8, B.1) compare two membership to show their to illustrate their purpose. With a triangular function we can see the membership rises linearly from `lo` directly to `hi` because `mid=hi`. Small deviations from baseline are already counted, which is useful when we want to accumulate weak but frequent shifts. The point near 4.0 has membership zero because it lies outside the `hi` range. With the one-sided exponential ramp, small deviations near `lo` contribute less and the curve rewards sustained or strong deviations. In practice this damps byte bursts caused by short background transfers while keeping the long or repeated nights prominent. A *healthy baseline* is defined shown as a green-band that ranges from `lo` to `mid`. Everything within this green-band won't indicate towards the assigned DSM-5 MDD criterion as we can see in both Figures 6.19 and 6.20. A transparent and controllable way to ignore an outlier and define the *healthy baseline*.

The calculated membership output time-series (Figure 6.17 reflects these choices: weeks of low membership are punctuated by a short period that crosses the threshold. We can see that also for a non-technical user like a parent or care-giver, this plot is easy to interpret, and it is directly linked to the raw plot and the membership curve, which keeps the workflow explainable.

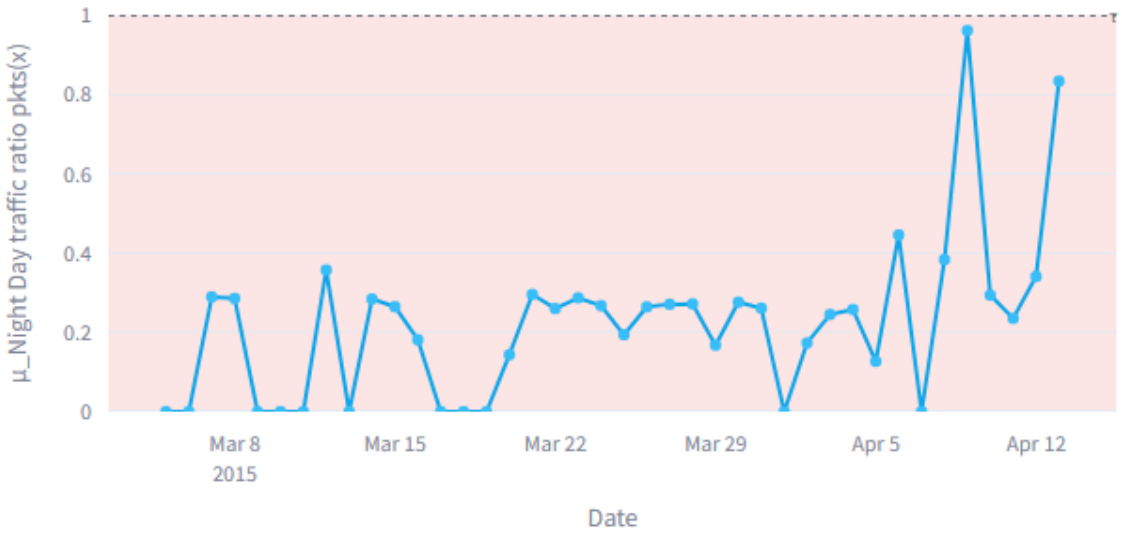


Figure 6.17: C4_F8 membership function output. Days within the red indicator band that exceed the threshold generate a clear signal between 0 and 1 for the DSM-5 MDD criterion gate.

6.2.2 Exposed Dataset Artifacts

The packet-count counterpart `C1_F8_NightDayRatioPkts` (B.1) behaves consistently with the byte-based ratio but is less sensitive to a few large transfers. The distribution in Figure 6.18 shows a baseline cluster well below 1.0 and several excursions toward 1.5. The membership panels in Figures 6.19 and 6.20 illustrate that both triangular and one-sided exponential shapes capture the same story but offer different control. With the triangular function, day-to-day changes near the baseline are emphasized. With the one-sided exponential, near-baseline noise is discounted and only sustained night-heavy periods accumulate membership. The combined reading is straightforward: when both `C4_F8` and `C1_F8` trend upward for several days, the evidence for a night-shift grows and when only bytes increase, we are likely seeing large downloads.

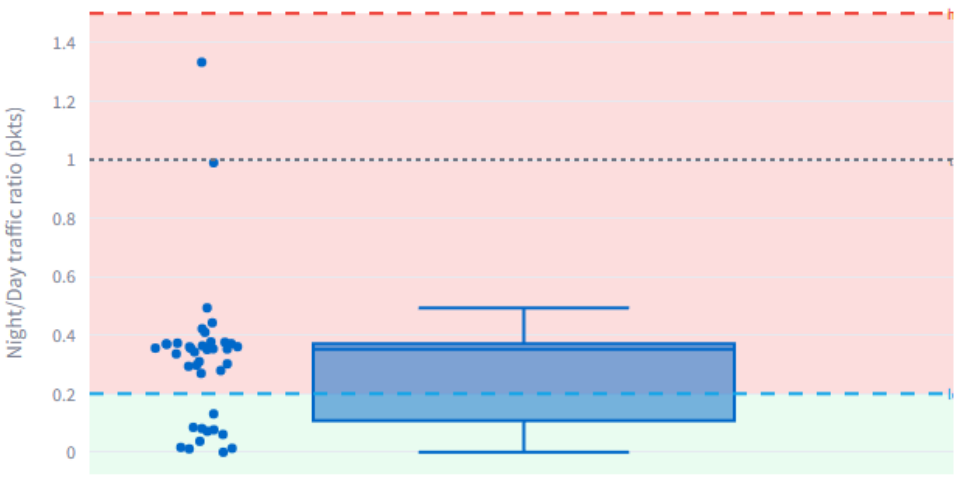


Figure 6.18: C1_F8 distribution. Most days fall below 1.0, with a small number of excursions that we can either accentuate or soften by the membership shape.

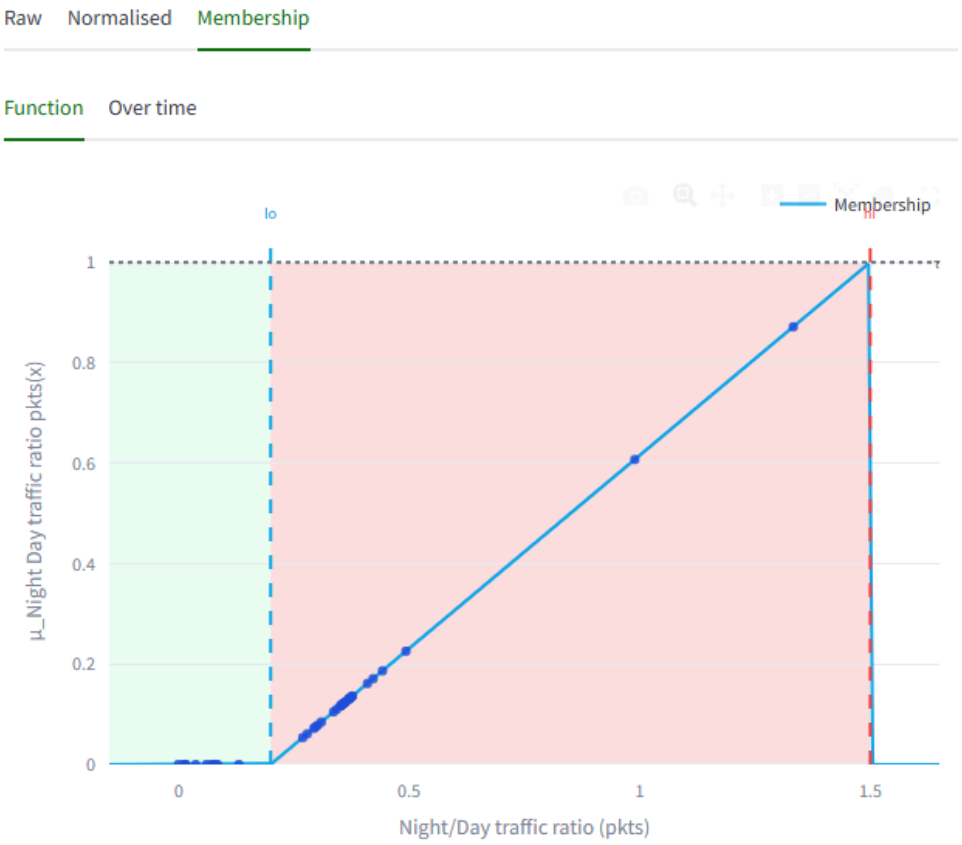


Figure 6.19: C1_F8 triangular membership. Linear increase from `lo` to `hi` highlights frequent but small shifts.

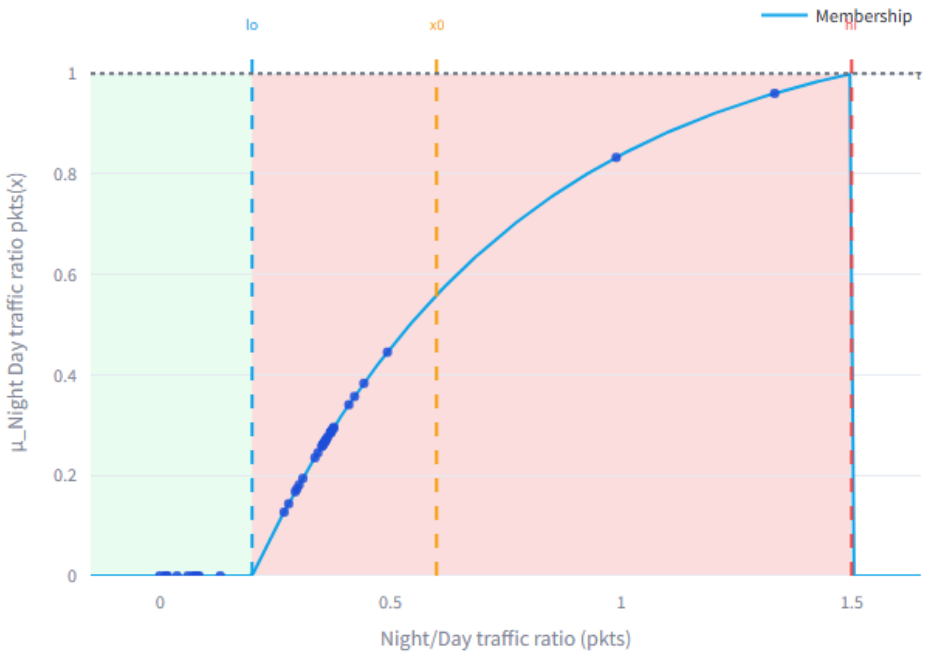


Figure 6.20: C1_F8 one-sided exponential membership. Near-baseline days contribute little, sustained deviations accumulate membership.

Figure 6.21 shows `C4_F2_WakeAfter0400Min`. In the workshop data most days report values near the upper bound with a single sharp dip. This is a realistic artifact of the source and can also be seen in the Figure 6.11 that shows continuous packet activity throughout the entire 24-hour period without major recording gaps. The continuous packet flow prevents the algorithm from detecting sleep onset and wake events, so the 'wake after 04:00' collapses to a constant. We can see that the implemented application makes such failure modes obvious in the raw over-time panel, which is valuable in practice. We down-weight the metric, keep a higher threshold, and rely more on the night/day ratios for this criterion.

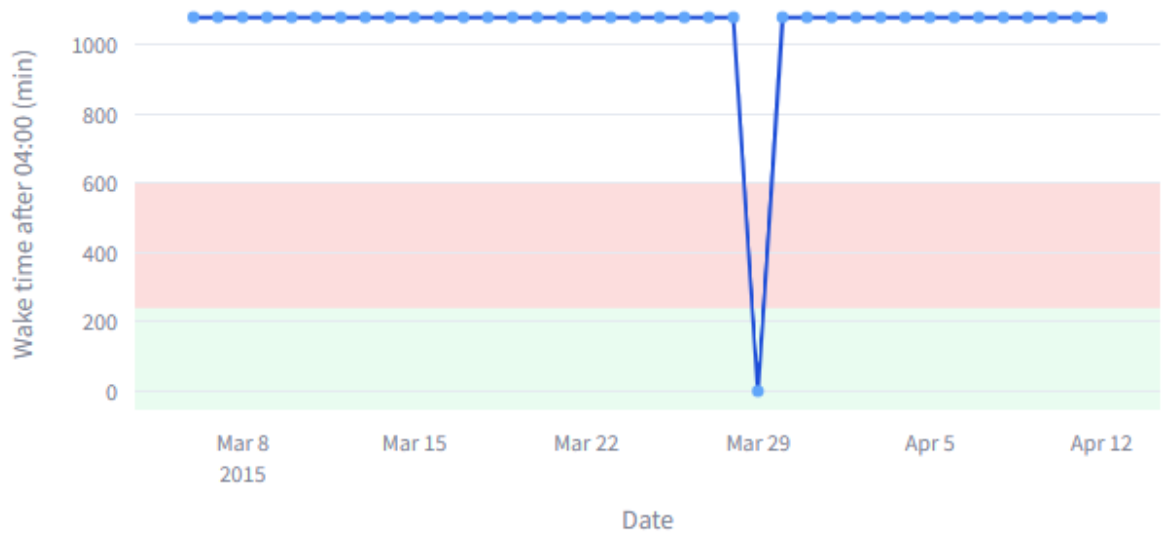


Figure 6.21: C4_F2 WakeAfter0400Min, raw input values. The near-constant trace with a single dip reveals a data-generation artifact in the workshop source. The UI makes the issue visible and easy to mitigate by down-weighting and thresholding.

6.2.3 Domain Diversity

We complement the above evaluated sleep-related ratios with two content-free, header-only indicators that speak to anhedonia. The number of distinct registrable domains visited per day (`C2_F1_UniqueSLD`, B.2) and the count of hits to a small set of productivity services (`C2_F6_ProductivityHits`, B.2). Both rely only on DNS names and HTTP host headers, so they preserve content privacy yet still expose changes in breadth of online activity and goal-directed behavior. In keeping with the workflow established above, we first read the over-time traces, confirm the baseline with the distribution view and then position `lo`/`mid`/`hi` and the membership shape so that routine days remain in green and only meaningful reductions raise membership.

Figure 6.22 shows the raw `C2_F1` time-series. Early in the period we see very broad exploration with many days between roughly one hundred and several hundred unique (Effective Top-Level Domain+1 (eTLD+1) domains. Around mid-March the pattern tightens and stays in a narrower band mostly between a few tens and about a hundred domains. This regime change is mirrored by the distribution in Figure 6.23, where the bulk of observations sits well above the red dashed boundary while a visible cluster approaches the lower end. Calibrating the triangular membership with `Invert` enabled and `lo`=20, `mid`=`hi`=40 treats sustained days with fewer than forty unique domains as

evidence for reduced exploration, while ignoring very large days as outliers above `hi`. The membership output over time in Figure 6.24 then highlights several short windows,late March to early April, where diversity remains compressed over consecutive days. This is exactly the kind of pattern that indicates towards the DMS-5 criterion for loss of interest/anhedonia. Not a single quiet day but a run of reduced variety compared to the persons own baseline. Because the limits are visible and editable, the user can always adapt them when weekends, travel or work sprints naturally reduce diversity for any user specific reasons.

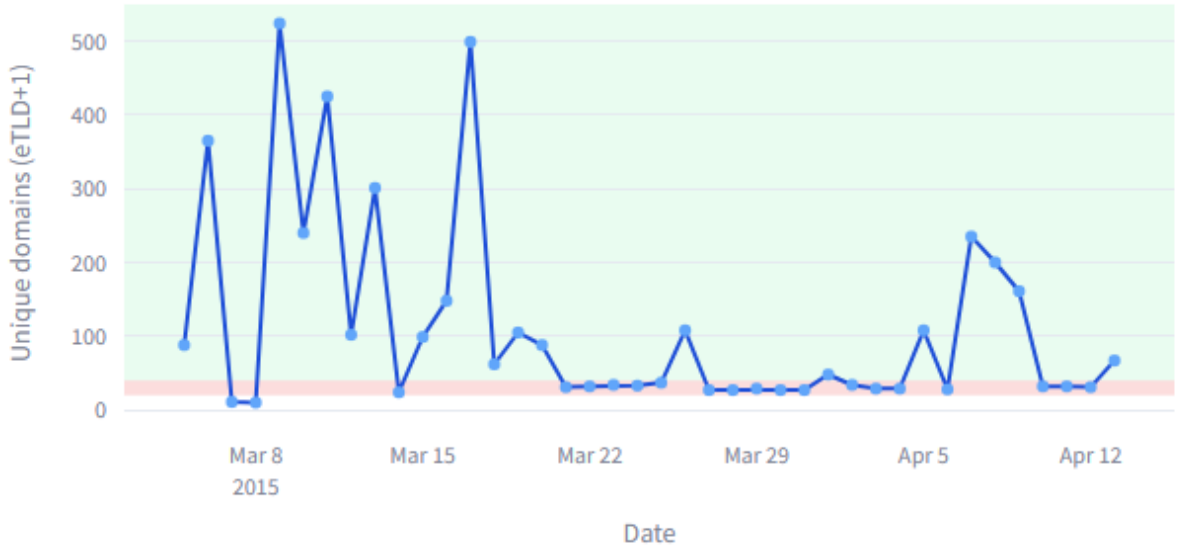


Figure 6.22: C2_F1 UniqueSLD. Broad exploration early in the period gives way to a narrower band; the green/red bands help set the baseline and indicator region.

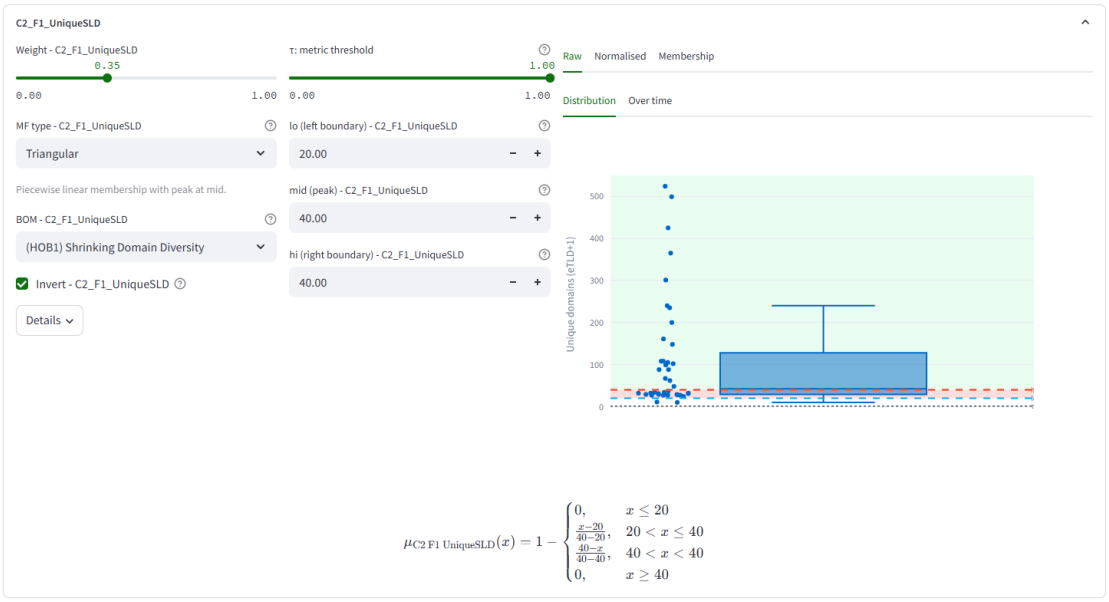


Figure 6.23: C2_F1 distribution. The box sits above the indicator boundary while several points approach the lower tail. This supports a left-shoulder calibration with `invert` so that lower diversity increases membership.

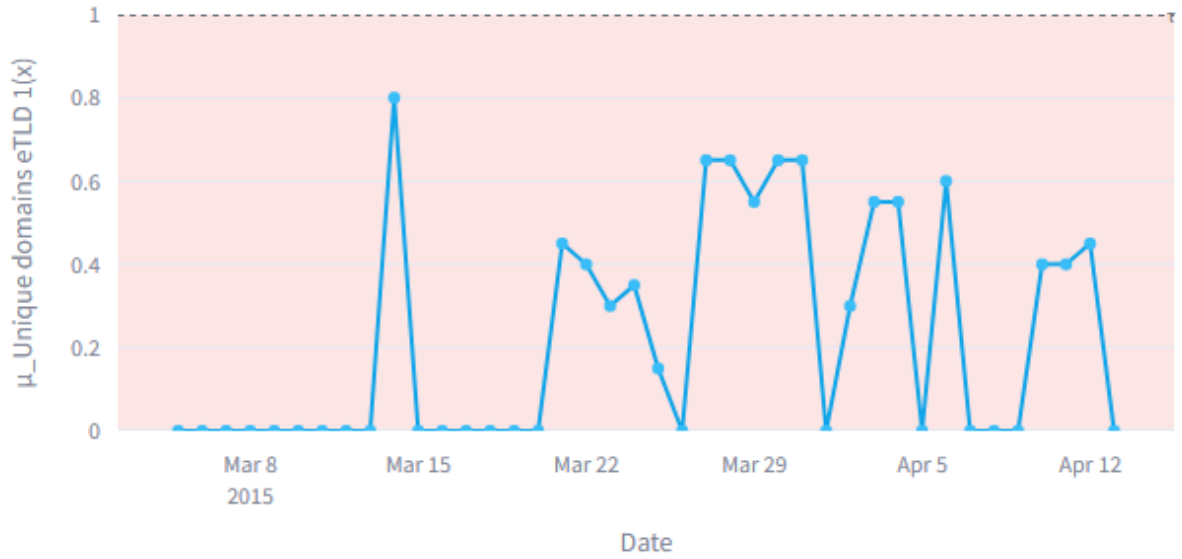


Figure 6.24: C2_F1 triangular inverted membership. Repeated reductions in domain diversity appear as runs of non-zero membership. Isolated large swings are ignored by construction because they lie above `hi`.

Figures 6.25 and 6.26 summarizes `C2_F6`. Most days fall between roughly 17'000 and 22'000 hits to the curated set. Only a few early days drop clearly below. With `invert` enabled and a triangular shape where `lo`=0 and `mid`=texttthi=15'000, the red band explicitly captures these unusually low-activity days while the bulk of typical days remains in green. The membership curve in Figure 6.27 then decays to zero as the count approaches the personal baseline. The resulting membership over time (Figure 6.28) shows a short early period with elevated membership followed by weeks at or near zero. This pattern is consistent with a possible brief phase of reduced planning/goal-oriented browsing that resolves-again, what we want the metric to DSM-5 MDD indication. Because the list of services is curated, users can adapt it to the persons toolset to avoid false negatives and because the function is left-shouldered and inverted, extremely high traffic on a single day will be ignored instead of being misinterpreted.

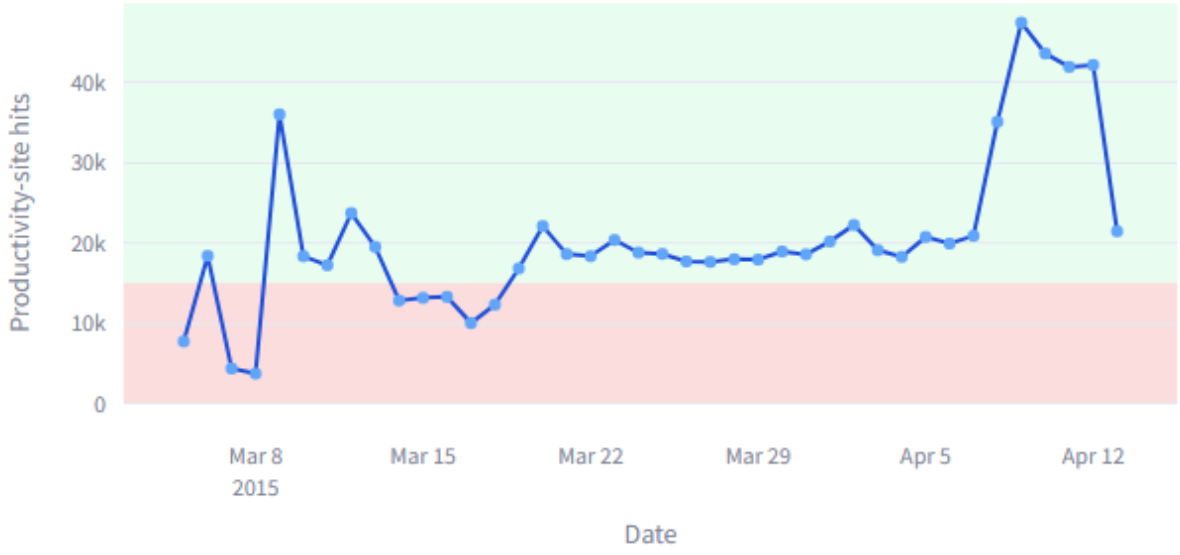


Figure 6.25: C2_F6 Productivity Hits. Fluctuations early in the period followed by a stable hit to 20'000 and ended with an increased productivity hit.

Figure 6.26: C2_F6 distribution. The central box indicates a stable baseline. Eight low days fall into the indicator band below 15'000 and are therefore treated as candidates for reduced goal-directed behavior.

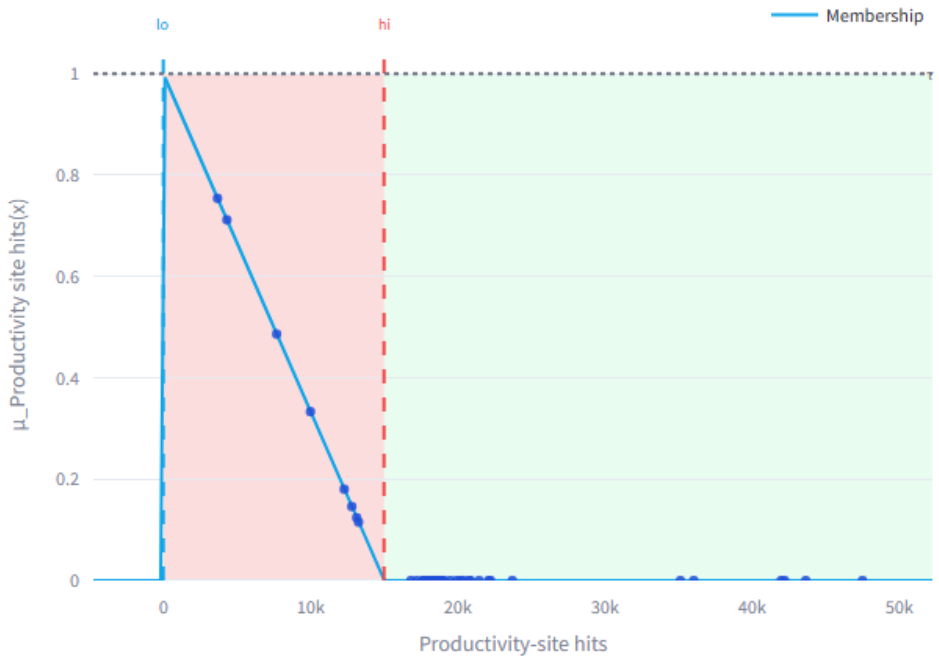


Figure 6.27: C2_F6 triangular and inverted membership function. Low counts map to high membership. Typical and high counts are in the green-band and contribute zero.

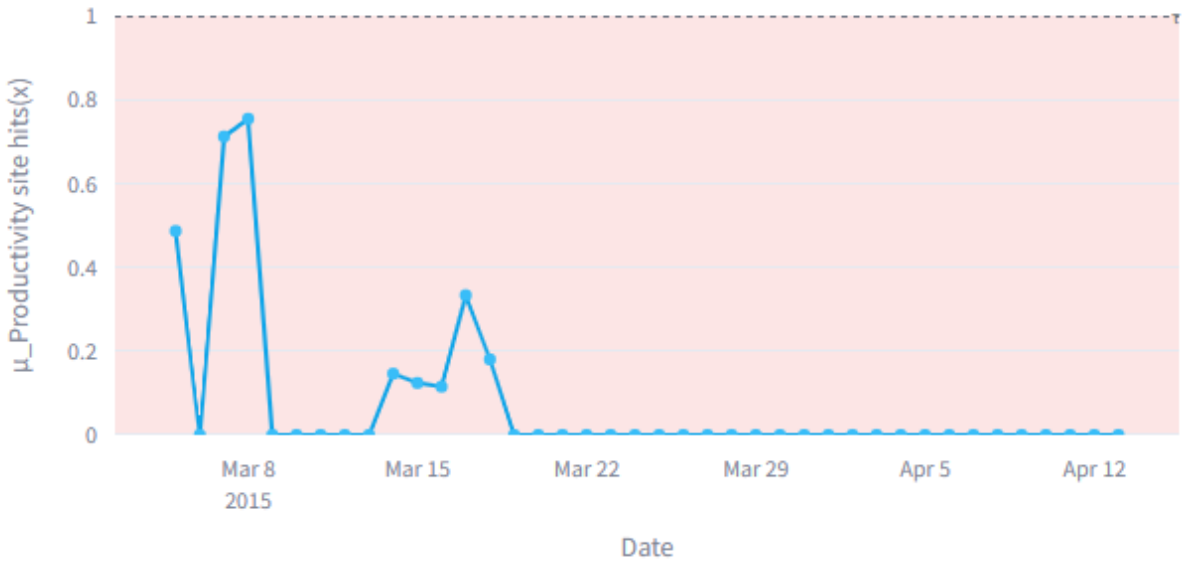


Figure 6.28: C2_F6 membership outputs. A brief early episode exceeds the indicator band and the rest of the period stays at defined baseline.

An improvement as a next step, picked up in the next Chapter 7, is that we plan to complement the current profile-aware manual parameter tuning with *explainable* automatic parameter setting (e.g., classic clustering-based, seasonality-aware auto-tuning of weights, `lo/mid/hi` and thresholds). Our initial implementation in this direction on Dataset B were not promising enough, likely due to input characteristics and limited time, so we defer a robust solution to future work.

6.2.4 Interpretable and Explainable Evidence

This subsection analyses the findings of our approach how five-minute packet partitions and DNS context are turned into day-level metrics, how we handle them with (`lo/mid/hi`, thresholds, membership shapes) and how the UI of our implemented app presents these choices in an explainable, auditable way. We use the figures to judge whether the method addresses the need for transparent features, person-specific baselines and robust thresholding across heterogeneous behaviors. We report what can be seen directly in the plots (trends, seasonality, outliers and their treatment), how these observations translate into criterion evidence. The limitations of this approach are collected at the end of this chapter and the discussion on the broader implications, if we closed the research gaps and where the improvements could be are discussed in the last Chapter 7.

The figures allow several evaluation statements about the system.
First, the combination of a distribution view and an over-time view is enough to set a credible baseline without long historical data. In our workflow we begin with the raw plot to see seasonality or regime changes, we confirm the baseline with the box plot, and we then position `lo`, `mid`, and `hi` so that healthy days collect in green and the indicator band starts where clinical meaning begins. The resulting membership curve is simple enough to explain to a non-technical user and the effect of every parameter move is immediately visible. This is a strong point of the implementation because it makes the method auditable.

Second, outlier management is tractable and transparent. Values above `hi` can be ignored by construction, which is exactly what we want when a single day shows a pathological ratio due to a one-off transfer. The triangular curve in Figure 6.15 and the point at 4.0 make this point tangible. There is no hidden clipping. The figure itself explains the decision.

Third, the system supports context that data alone does not capture. If a user reports a seasonal condition, for example hay fever that reduces daytime device use for a week, the analyst can either reduce the weight of the affected metric or shift the baseline by increasing `lo`. Both actions are visible in the membership plot and reversible. This is preferable to hard coding exceptions because it keeps the model simple and the reasoning documented next to the figure.

Fourth, the approach scales across metrics. We applied the same visual language and the same controls to byte ratios, packet ratios and time-of-day markers. The resulting memberships are directly comparable and can be combined within a criterion. In other words, once a users learns to read one metric and its figures, they can read them all.

Fifth, the user can choose between sensitivity profiles rather than a single fixed rule. Triangular functions accumulate many small deviations whereas one-sided exponentials suppress noise and emphasize sustained change. The choice affects only how fast the membership rises, not the semantics of the metric which remains e.g. 'more night than day' or 'more packets at night than during the day'. This separation of semantics and sen-

sitivity is a design strength and can also be scaled with additonal membership functions.

In addition, the anhedonia-oriented metrics `C2_F1_UniqueSLD` and `C2_F6_ProductivityHits` reinforce these findings with content-free signals derived from packet headers only. The `UniqueSLD` panels (6.22, 6.23, 6.24) show that we can read meaningful contractions in the breadth of online activity over consecutive days and the membership plot makes the decision line explicit: a person-specific baseline is set by `lo`/`mid`/`hi`, outliers above `hi` are harmless by construction and the `invert` option matches the semantics that 'less indicates to criterion' without redefining the metric. The `ProductivityHits` panels (6.26, 6.27, 6.28) capture a complementary aspect of goal-directed behavior. A short early episode falls into the indicator band while the following weeks remain at baseline. Both examples confirm that our tuning knobs generalize beyond sleep. They are explainable at a glance and robust to benign large days because the indicator band is bounded. Even though Dataset B is not a record of a single individuals everyday life, the method still surfaces plausible episodes (reduced domain diversity in late March, possible low productivity counts early on) and keeps them interpretable. We therefore strengthen the claim that a router-centric, membership-based approach can bridge from low-level signals to DSM-5-relevant indicators while remaining transparent about uncertainty. The UI makes this explicit and offers two levers. Adjusting the limits and re-weighting the metric, whenever context (for example vacations, illness, or seasonal allergies) suggests a temporary shift that should not over-influence a criterion.

Finally, the evaluation exposes a real limitation in the Swedish workshop dataset that our tool helps to handle. Always-on traffic reduces the value of sleep-like metrics such as e.g. `C4_F2`. Because the plots make that visible, we can mitigate it by down-weighting and by leaning on more robust metrics in the same criterion.
An improvement as a next step, picked up in the next Chapter 7, is that we plan to complement the current profile-aware manual parameter tuning with *explainable* automatic parameter setting (e.g., classic clustering-based, seasonality-aware auto-tuning of weights, `lo`/`mid`/`hi` and thresholds). Our initial implementation in this direction on Dataset B were not promising enough likely due to input characteristics and limited time, so we defer a robust solution to future work. A detailed reflection on how these parameters were chosen and how they influence our results follows in the next Section 6.2.5.

6.2.5 Parameter Setting and its Influence

Before aggregating the tuned metrics into criterion-level likelihoods, it is important to reflect on how the parameters for each metric were defined and what implications they have for the final indication results. The choice of `lo`, `mid`, `hi`, as well as the weights and thresholds, was and will not be a purely objective process. Each parameter was set through iterative inspection of the plots presented earlier and by combining statistical tendencies observed in Dataset B with contextual understanding of what each metric is intended to represent. In other words, the calibration relied on both data-driven evidence and domain reasoning, when a metrics interpretation suggested that 'more' or 'less' activity could indicate a DSM-5 MDD symptom, the corresponding membership function and its inversion were adjusted accordingly. This process can therefore be seen as a best-effort approximation of realistic parameter values under the data constraints available. Because Dataset B does not reflect authentic multi-user behavior and contains continuous network activity, the healthy baseline and indicator bands for many metrics had to be estimated conservatively rather than empirically confirmed. The chosen limits aim to capture plausible transitions between normal and abnormal ranges, but they remain flexible assumptions rather than fixed thresholds. Their selection has a direct and sometimes substantial impact on the resulting criterion likelihoods evaluated in Section 6.3, especially when several metrics reinforce each other or when a single highly weighted metric changes its slope.

For this reason, we can highlight again that also for parameter setting transparency and reproducibility are central to our implementation. All values are stored and can be adjusted in the app, allowing users or clinicians to rerun the evaluation with modified assumptions. The broader methodological implications of these parameter dependencies, and how they might be mitigated by automated or user-specific tuning approaches, will be discussed further in Chapter 7.

With this, we now proceed to the criterion-level evaluation and aggregate multiple tuned metrics per DSM-5 criterion into a day-level likelihood using the weighting, thresholds and gate logic defined in Chapter 4. This follows the idea of aggregation introduced in system design and allows us to study how corroborating metrics change the criterion signal over time.

6.3 Aggregation

We now turn from per–metric calibration to DSM-5-criterion–level evidence. For each DSM-5 criterion we aggregate some selected tuned metrics into day–level likelihoods using the weighting, thresholds, and gate logic defined in Chapter 4 and evaluated before in Section 6.2. The aggregation follows the design idea already introduced above where we first compute per–metric memberships $\mu_m(x_m)$, then combine them as a weighted score. The resulting likelihoods are shown both as a contribution trace (which metric added how much on a given day) and as a criterion curve over time with a dashed decision line acts as the DSM-5 gate at $\theta = 0.6$. Parameters for all metrics in this section are the an asumption and based on our analysis of Dataset B. The parameters can be stored in the application configuration, therefore every result evaluated in this section is reproducible within the app [36].

6.3.1 C1 Depressed mood

Table 6.2 lists the metrics chosen for Criterion 1 together with their weights and membership limits. The direction indicated by "↑→evidence" reflects the current calibration (i.e., whether higher values after `lo` increase membership).

Metric (short rationale)	w	MF	lo	mid	hi	dir.
C1_F1_DistinctSocial - More distinct social domains than usual may reflect restless, fragmented social activity; can be inverted to capture withdrawal if desired.	0.25	tri	20	80	80	↑
C1_F4_DownUpRatio - Large down/up ratios indicate passive consumption over active posting, which aligns with reduced initiative.	0.20	tri	3	12	12	↑
C1_F6_Fano - Bursty minute-level activity (high Fano) suggests irregular, stop-and-go usage; we use it as a weak indicator of instability.	0.10	tri	100	1000	1000	↑
C1_F7_HourlyCV - Concentration of activity into a few hours (high Coefficient of variation) can accompany unregular daily structure.	0.10	tri	0.6	3.0	3.0	↑
C1_F8_NightDayRatioPkts - Shift of packet activity into the night, consistent with late-night wakefulness.	0.35	tri	0.2	1.0	1.0	↑
dataset/ datset group = Dataset B / [ALL_OTHER]						
M=14 (rolling window, days) / N=6 (days $\geq \theta$)						
$\theta = 0.6$ / $\tau = 1$ (criterion and metric threshold)						

Table 6.2: DSM-5 MDD Criterion 1 metric configuration (weights and triangular MF limits) and their intended contribution to depressed mood.

Figure 6.29 visualizes the information flow from PCAP windows to C1 features (F1, F4, F6, F7, F8) and their parameters used for the calculation to the DSM-5 MDD C1 indicator.

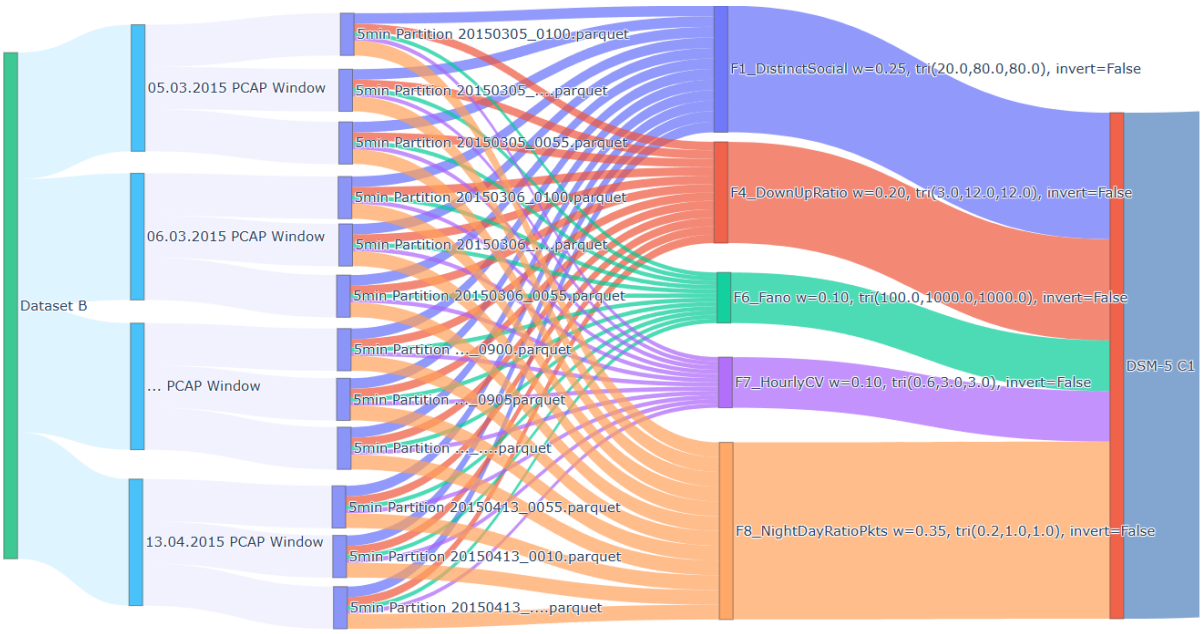


Figure 6.29: Criterion C1 (Depressed mood) Information Flow. Router-level daily PCAP windows are split into 5-minute partitions, transformed into the configured C1 features, and aggregated into the DSM-5 MDD C1 likelihood.

Figure 6.30 shows the weighted membership contributions by day. The pattern matches what the underlying metric plots suggested, but the stacked view also reveals how the metrics interact. C1_F8 (night/day packets, $w = 0.35$) provides the backbone of the signal by marking clear shifts of traffic into typical sleep hours. On its own it can raise the criterion but cannot cross the gate, which is intentional so that a single nocturnal episode does not dominate C1. On days where C1_F1 (distinct social domains, $w = 0.25$) rises together with C1_F4 (down/up ratio, $w = 0.20$), the curve reflects a change in interaction style. Broader and more restless social footprints combined with more passive consumption add secondary weight. Isolated spikes of C1_F7 (hourly CV, $w = 0.10$) and C1_F6 (Fano, $w = 0.10$) capture within-day concentration and burstiness. When these co-occur with night-heavy activity, we observe the moderate composite peaks around 0.35–0.45 in the likelihood trace. When they do not align temporally, the score remains low, which is exactly the behavior we want from a corroboration scheme: complementary axes, circadian shift (C1_F8), breadth and style of social engagement (C1_F1, C1_F4), and within-day structure (C1_F7, C1_F6) must reinforce each other to suggest depressed mood. This interplay could keep false positives down while preserving sensitivity to short multi-metric episodes. In future iterations the same mechanism can be strengthened with additional household-level signals to improve coverage, which we outline in Chapter 7.

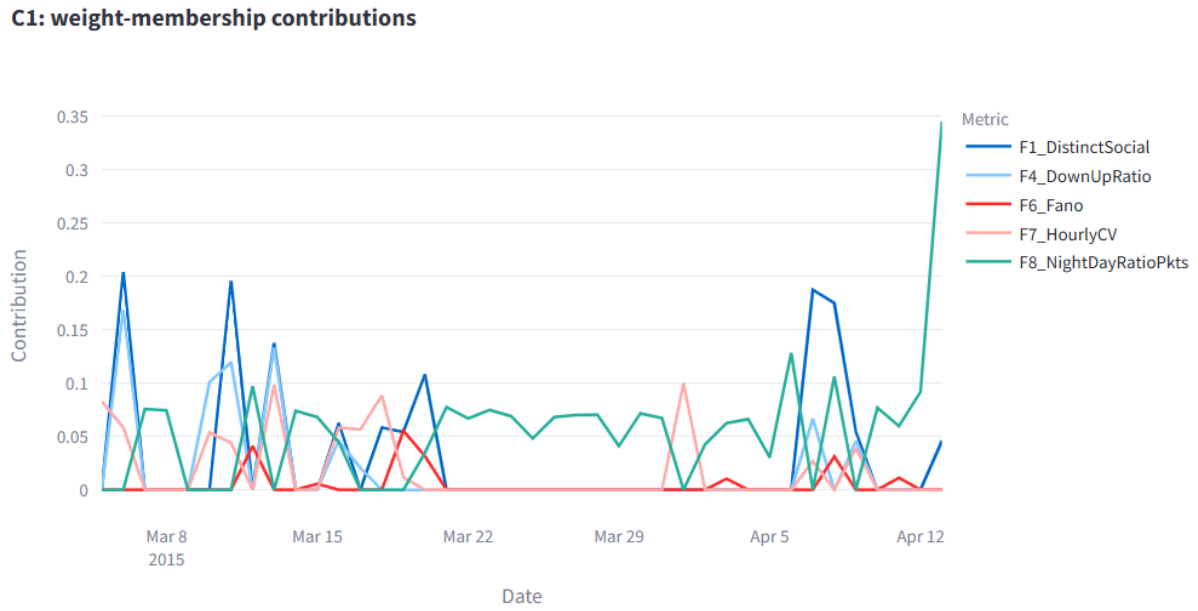


Figure 6.30: C1: weight–membership contributions over time. C1_F8 drives the late spike. Earlier contributions come from C1_F1 and C1_F4 with occasional C1_F7/F6 peaks.

The aggregated likelihood curve in Figure 6.31 stays in the green band for most of the 40 days, with short excursions to roughly 0.35–0.45 early in March and again near the end. The dashed DSM-5 gate at $\theta = 0.6$ remains above the observed values, so the criterion gate is not triggered in this run. This is broadly what we expect from Dataset B: a few nights with pronounced nocturnal activity and some days with more passive consumption or broader social footprints, but no sustained multi-metric pattern crossing the gate.

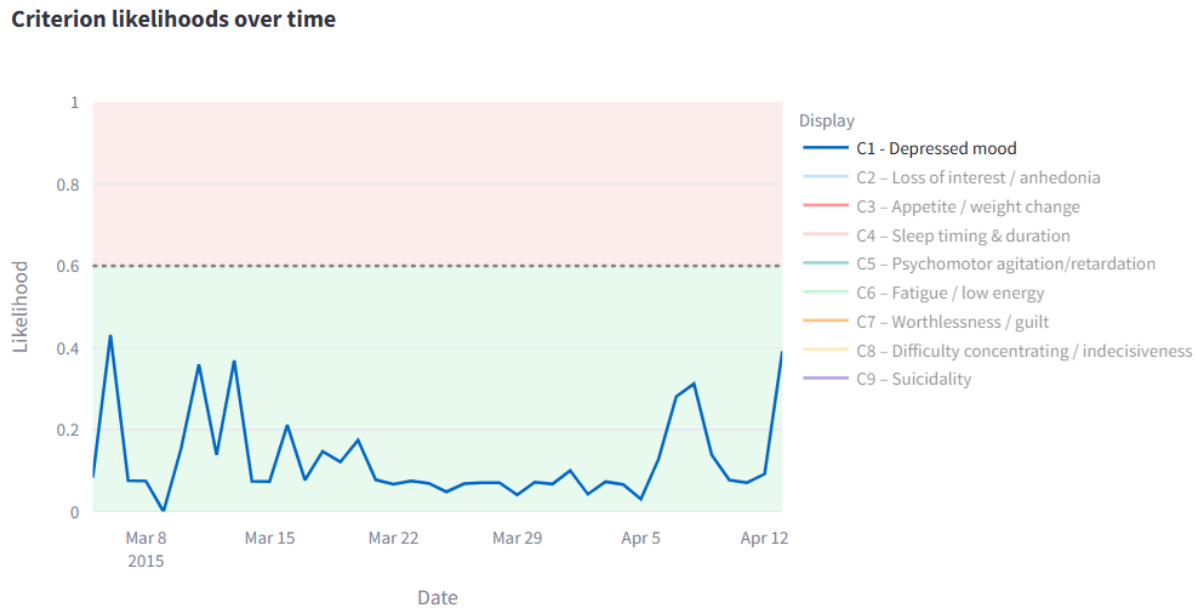


Figure 6.31: C1 criterion likelihood over time. The likelihood remains below the decision line at $\theta = 0.6$, with short, explainable rises.

Figure 6.32 summarizes the 40-day average at 0.127. This sits clearly in the green range and is consistent with the time-series above.

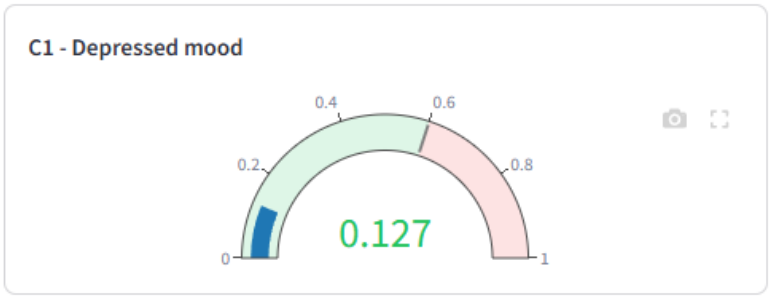


Figure 6.32: C1: 40-day average likelihood. An easily readable entry-point indicator for care-givers or parents, here the mean likelihood is rather low with 0.127.

6.3.2 C2 Loss of interest / anhedonia

Table 6.3 lists the metrics we selected for this evaluation for Criterion 2 together with the weights and membership limits used in our run. For this criterion lower values after `lo` are meant to increase membership, therefore all four metrics are calibrated with the invert option. The settings represent our best assumption from Dataset B and the metric semantics, and they are stored in the configuration so that the results reported here can be reproduced with the app.

Metric (short rationale)	w	MF	lo	mid	hi	dir.
C2_F1_UniqueSLD - Fewer distinct registrable domains over several days suggest a narrowing of online exploration that is consistent with reduced interest.	0.40	tri	0	40	40	↓
C2_F3_ChatSessionCount - Fewer messaging sessions indicate fewer spontaneous or planned conversations and can reflect reduced social drive.	0.05	tri	1	3	3	↓
C2_F4_MeanUpstreamRateBps - Lower average upstream rate during chat points to lurking rather than active posting which fits anhedonia.	0.25	tri	0	100	100	↓
C2_F6_ProductivityHits - Fewer visits to calendars notes and task tools point to less planning and follow-through.	0.30	tri	3500	15000	15000	↓
dataset/ datset group = Dataset B / [ALL_OTHER]						
M=14 (rolling window, days) / N=6 (days $\geq \theta$)						
$\theta = 0.6$ / $\tau = 1$ (criterion and metric threshold)						

Table 6.3: DSM-5 Criterion 2 metric configuration and intended contribution to loss of interest / anhedonia.

Figure 6.33 shows how the information flow to gt domain diversity, messaging activity, upstream posting, and productivity signals combined into the C2 indicator.

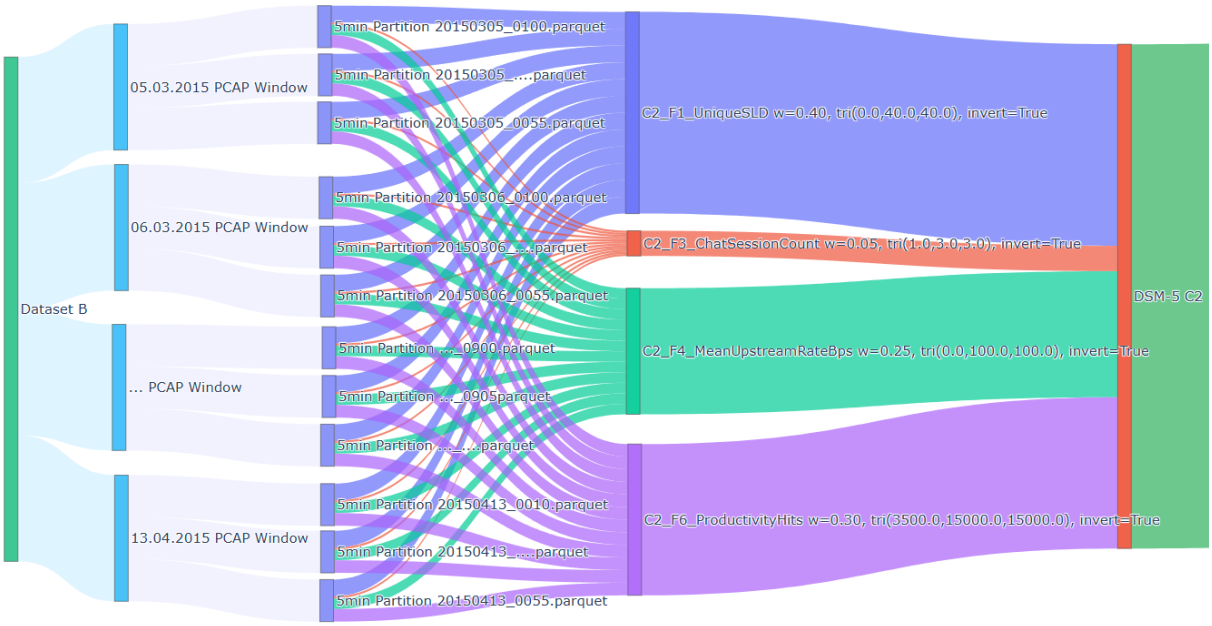


Figure 6.33: Criterion C2 (Loss of interest / anhedonia) Information Flow. Daily PCAP windows feed C2 feature extraction (e.g., UniqueSLD, ChatSessionCount, MeanUpstreamRate, ProductivityHits) whose parametrized, weighted memberships compose the DSM-5 MDD C2 score.

We now look at how these in Table 6.3 mentioned metrics combine over time. Figure 6.34 shows the weighted membership contributions for each day. The first week contains a short episode in which ProductivityHits and MeanUpstreamRate both drop and therefore contribute strongly because both are inverted. At the same time UniqueSLD also contributes because domain diversity is low on those days. The three signals reinforce one another and produce the highest composite contributions in the entire period. Later in March and April the picture changes. UniqueSLD becomes the main source of small but recurring contributions that reflect a milder narrowing of day-to-day variety, while ChatSessionCount provides a low and steady background contribution due to its small weight and inverted calibration. The contribution plot therefore shows a convincing interaction. Large reductions in goal-oriented activity and active posting create sharp early peaks. After that the criterion is carried by gentle evidence from reduced variety with only minor support from messaging volume. This interaction is exactly what we intended when we combined the features. One group captures engagement with planning tools and active posting, another group captures breadth of exploration. When both groups move together we see a pronounced episode. When only one group moves we see a low-level signal that is still meaningful but does not dominate the criterion.

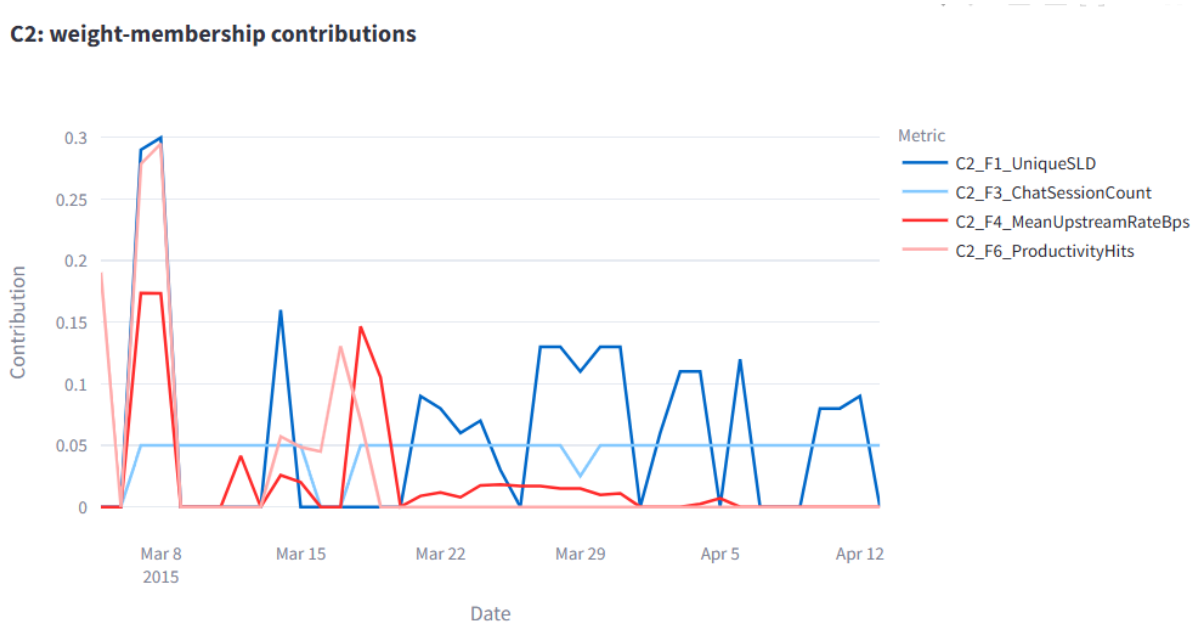


Figure 6.34: C2 weight–membership contributions over time. Early peaks arise when several inverted metrics drop simultaneously. Later contributions are mainly driven by UniqueSLD with a small steady background from ChatSessionCount.

The aggregated likelihood curve in Figure 6.35 reflects this story. There is a short rise above 0.8 in the first week where several metrics align, followed by a long period in the green band between about 0.05 and 0.30 with occasional pulses when UniqueSLD and either ProductivityHits or UpstreamRate dip together. The dashed gate at $\theta = 0.6$ is crossed only during the early episode, which matches the contribution view and what we saw in the metric panels. This is a positive result for explainability. The criterion rises when multiple aspects of interest and initiative decrease together and it stays low when only one aspect fluctuates.

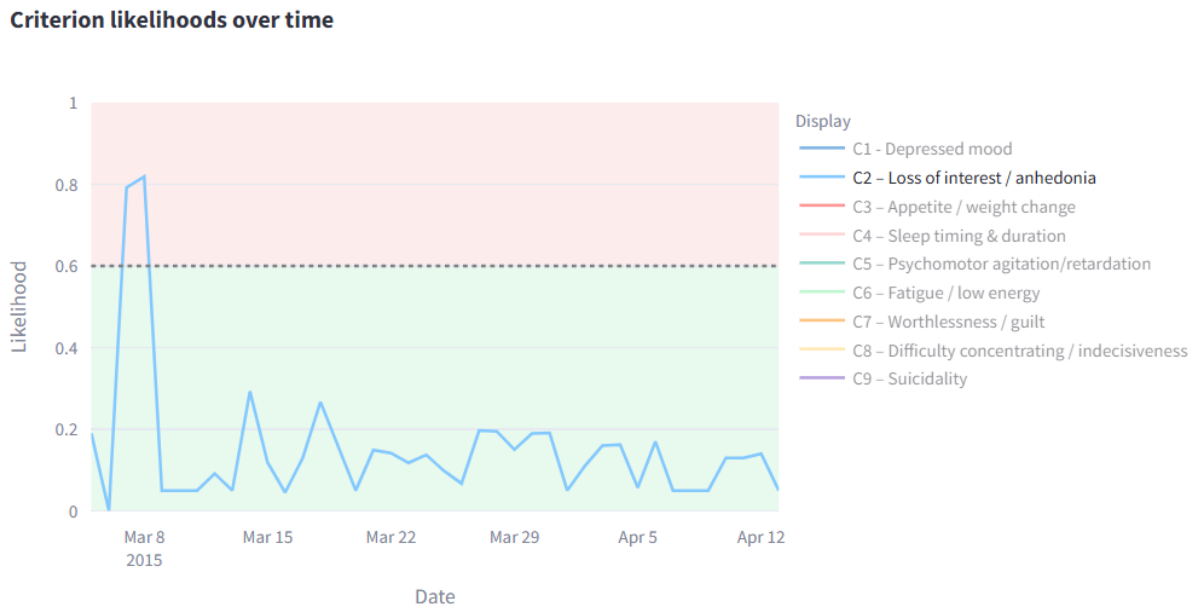


Figure 6.35: C2 criterion likelihood over time. A brief early episode exceeds the gate; the remainder stays below the threshold with small, explainable pulses.

The gauge in Figure 6.36 summarizes the 40-day average at 0.152. This is consistent with the time-series and provides an easy entry point for a clinician or care-giver. It encourages a closer look at the early days and at weeks where domain diversity compresses without major changes in planning or posting.

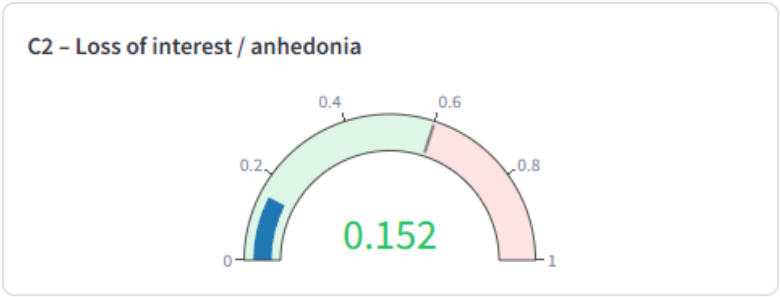


Figure 6.36: C2 40-day average likelihood. The mean remains low which matches the predominately green period after the early episode.

6.3.3 C4 Sleep timing and duration

Table 6.4 summarizes the metrics we use within this evaluation for the sleep criterion together with the weights and membership limits in this run. The calibration places more weight on late wake after 04:00 and adds complementary signals from sleep duration deviations, night–day traffic balance and daytime idle structure. The settings are our best assumption from Dataset B and from the intended meaning of each metric; they are stored in the configuration so that every result here can be reproduced with the app.

Metric (short rationale)	w	MF	lo	mid	hi	dir.
C4_F2_WakeAfter0400Min - Later wake after 04:00 increases membership and marks delayed sleep timing when it persists across days.	0.65	tri	120	1085	1085	↑
C4_F4_SleepDurationZAbs30d - Large absolute z-scores of sleep duration relative to the 30-day baseline suggest atypical sleep length.	0.20	tri	0.25	0.80	0.80	↑
C4_F7_DaytimeIdleRatio0818 - Mid-range daytime idle ratio points to fragmented daytime structure which often co-occurs with irregular sleep.	0.05	tri	0.00	0.08	0.16	↑
C4_F8_NightDayTrafficRatioBytes - More traffic at night than during the day indicates night-time wakefulness.	0.15	tri	0.20	1.00	1.00	↑
dataset/ datset group = Dataset B / [ALL_OTHER]						
M=14 (rolling window, days) / N=6 (days $\geq \theta$)						
$\theta = 0.6$ / $\tau = 1$ (criterion and metric threshold)						

Table 6.4: DSM-5 Criterion 4 metric configuration and intended contribution to sleep timing and duration.

The information flow in Figure 6.37 clarifies that late wake, duration deviations, daytime idle structure, and night-heavy traffic jointly determine the C4 likelihood.

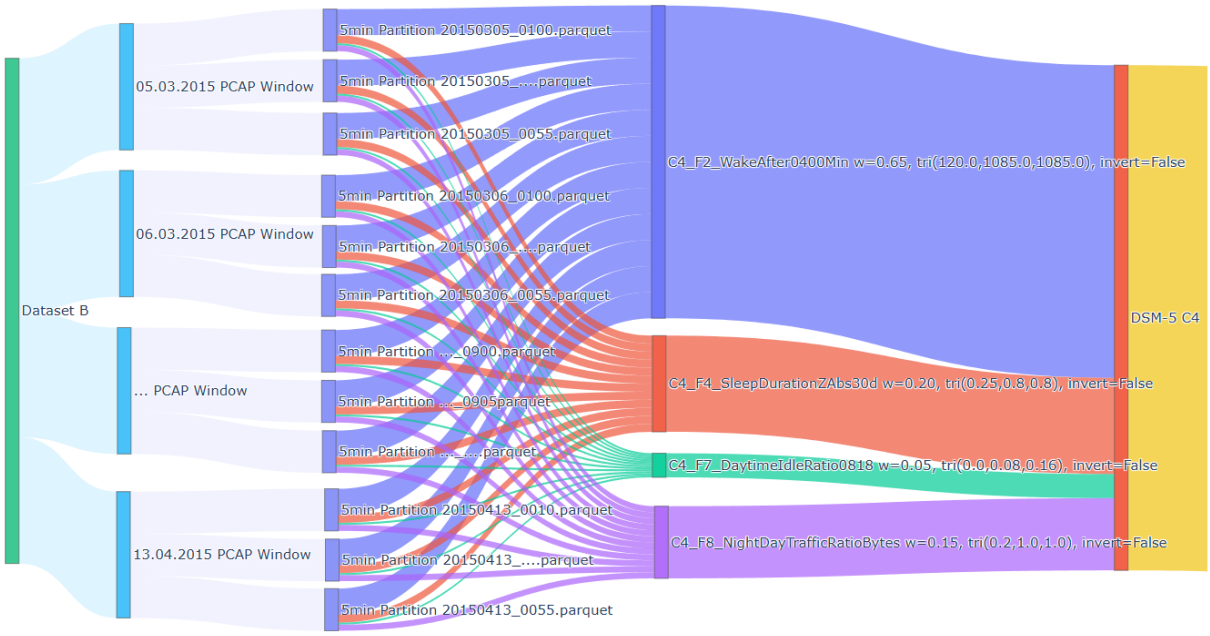


Figure 6.37: Criterion C4 (Sleep timing and duration) Information Flow. From daily PCAP windows to sleep-related features (WakeAfter0400Min, SleepDurationZAbs30d, DaytimeIdleRatio, Night/Day traffic ratio) and their aggregation into DSM-5 MDD C4.

We now examine how these metrics interact over time. Figure 6.38 displays the weighted membership contributions. C4_F2 carries most of the score because its weight is high and its membership remains near one on nearly all days. This creates a visible plateau that only breaks for a single day around the end of March. The application makes this behavior easy to diagnose. In Dataset B the continuous packet flow often prevents clear sleep events so the wake-after-04:00 estimator saturates. Leaving the settings untouched here is useful as it demonstrates how the contribution view immediately reveals that one metric dominates the criterion. The remaining features still play an informative role. C4_F4 shows short early transients that indicate sleep duration deviating from the rolling baseline. C4_F8 adds small pulses when the night–day balance tilts toward night-time. C4_F7 rises a little toward the end which is consistent with brief increases in daytime idle structure. Together these signals tell a coherent story. When C4_F2 is high and one of the other features moves in the same direction the daily score increases above the plateau, and when C4_F2 dips the whole criterion resets to a very low value, which we see on the single day with a near zero contribution.

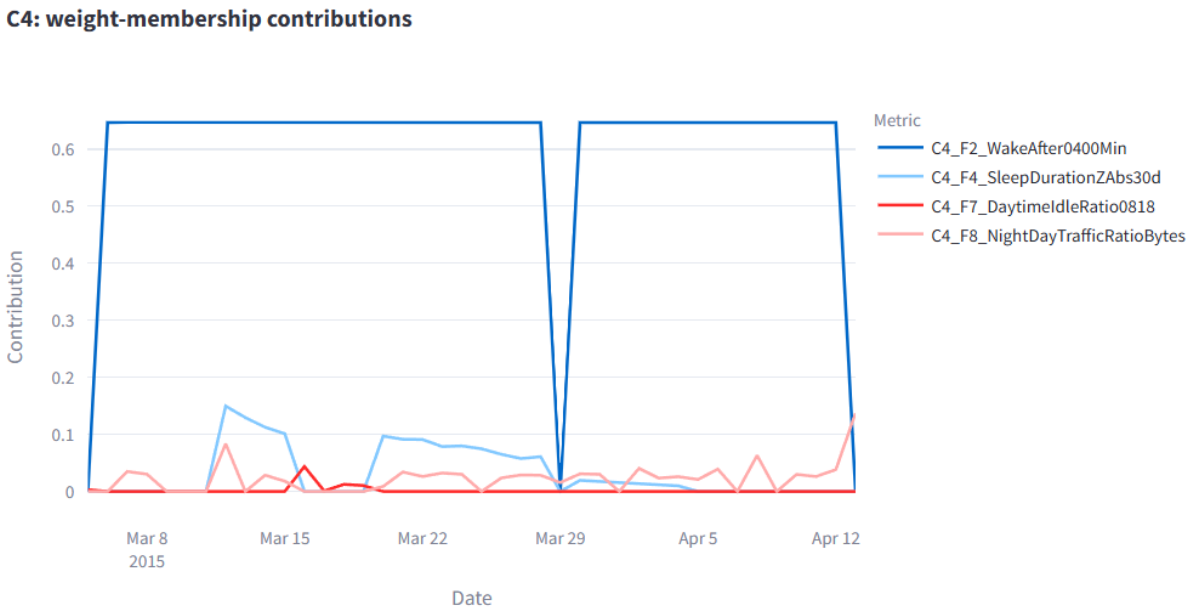


Figure 6.38: C4 weight–membership contributions over time. C4_F2 dominates most days while C4_F4, C4_F8 and C4_F7 add short pulses that refine the picture.

The aggregated likelihood curve in Figure 6.39 mirrors the contributions. The curve stays above the decision line for long stretches with values around 0.65 to 0.9 and it briefly drops to nearly zero on the day where the wake-after-04:00 signal collapses. This is consistent with the known characteristics of Dataset B and again shows the value of the workflow. The plots make the saturation visible and the parameter panel allows straightforward mitigation by down-weighting C4_F2 or by adjusting its threshold if we want to rely more on the corroborating metrics. We kept the configuration here to emphasise how the UI supports this kind of reasoning.

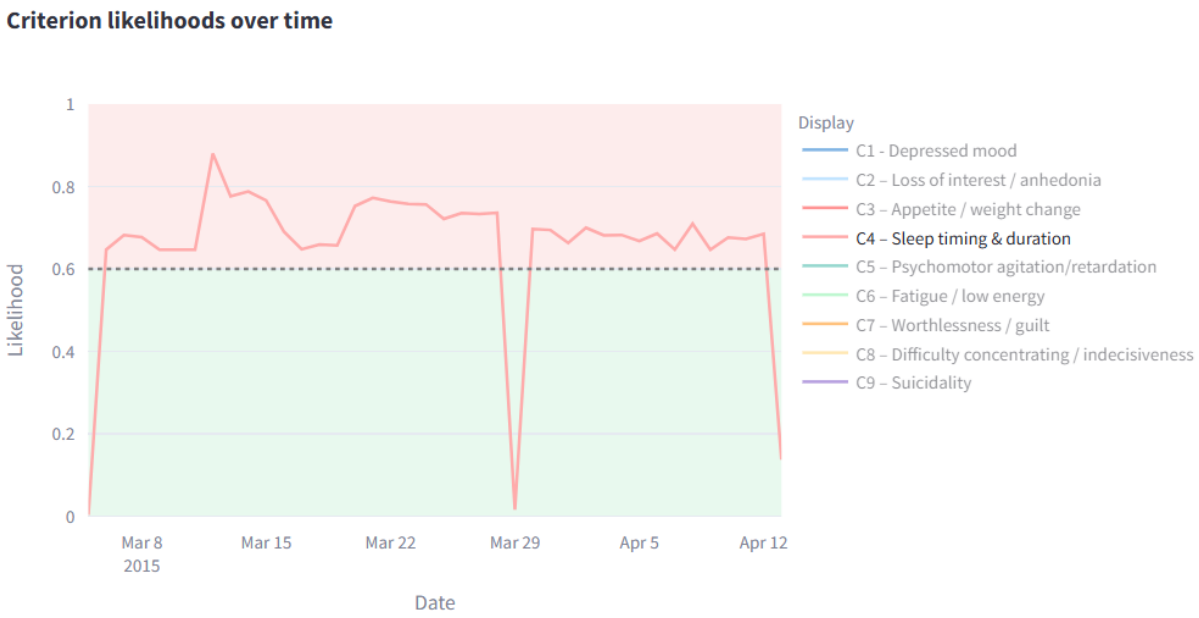


Figure 6.39: C4 criterion likelihood over time. Long stretches above the gate reflect a saturated wake-after-04:00 signal with small refinements from the other metrics and a single reset day near the end of March.

The gauge in Figure 6.40 summarizes the 40-day average at 0.655 which sits just above the gate and matches the time-series. For a clinician or care-giver this acts as an immediate entry point suggesting that sleep timing and duration deserve a closer look. The contribution view then helps decide whether the high score is driven mainly by late wake detection or whether night-time activity and duration deviations reinforce it. The same aggregation scheme is flexible and can be enriched with additional sleep-related features if the household-level metric set would be expanded with other digital environmental signals, which we discuss in Chapter 7.

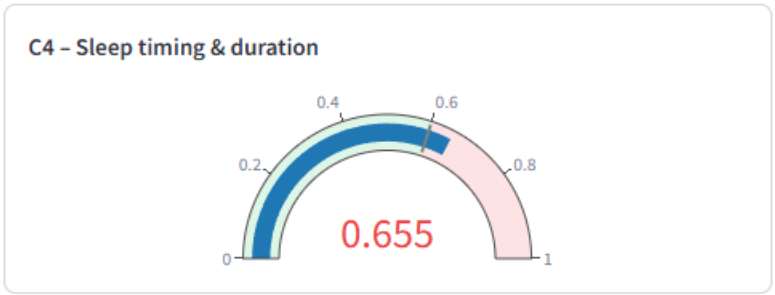


Figure 6.40: C4 40-day average likelihood. The mean is above the decision line which invites further inspection of the underlying sleep metrics.

6.3.4 C8 Difficulty concentrating / indecisiveness

Table 6.5 lists the metrics and settings we use for this DSM-5 MDD criterion 8. The focus is on content-free indicators of attention stability at the DNS and typing level. All three metrics are calibrated so that higher values after `lo` increase membership. Here again, the configuration represents our best assumption from Dataset B and from the intended meaning of each metric. The exact parameters can be configured within the application and the plots below can be reproduced.

Metric (short rationale)	w	MF	lo	mid	hi	dir.
C8_F2_DNSBurstRatePerHour - Frequent bursts where several different sites are contacted within a minute suggest quick hopping between topics and therefore less steady focus.	0.40	tri	25	61	61	↑
C8_F4_RepeatedQueryRatio60m - High ratios of identical queries within an hour can mean repeated checking or forgetting, which aligns with indecision.	0.40	tri	0.80	1.00	1.00	↑
C8_F8_MedianIKSsec - Longer median inter-keystroke gaps point to slower, less fluent typing and can accompany lapses in concentration.	0.20	tri	0.12	0.22	0.22	↑
dataset/ datset group = Dataset B / [ALL_OTHER]						
M=14 (rolling window, days) / N=6 (days $\geq \theta$)						
$\theta = 0.6$ / $\tau = 1$ (criterion and metric threshold)						

Table 6.5: DSM-5 Criterion 8 metric configuration and intended contribution to difficulty concentrating or indecisiveness.

Information flow in shown in Figure 6.41 depicts how bursty DNS activity, repeated lookups, and typing dynamics corroborate the C8 indicator.

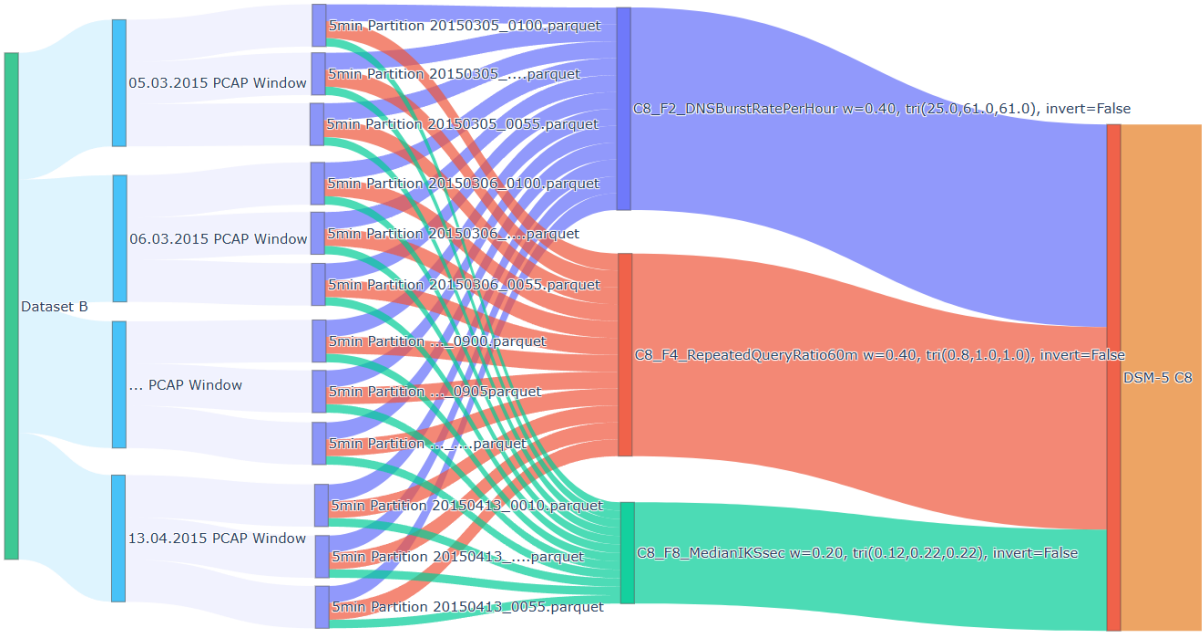


Figure 6.41: Criterion C8 (Difficulty concentrating / indecisiveness) Information Flow. Daily PCAP windows are partitioned and mapped to attention-stability proxies (DNS burst rate, repeated-query ratio, median inter-keystroke gap) that form the DSM-5 MDD C8 score.

We now consider how the three metrics combine over time. Figure 6.42 shows the weighted membership contributions for each day. Early in the period the repeated-query ratio contributes most and forms tall light-blue columns. These days show many within-hour retries and the membership function translates that into strong evidence. From the third week onward the DNS burst rate becomes the main driver and creates a stable dark-blue plateau around 0.35 to 0.40. This indicates many short sequences of distinct sites contacted in quick succession and it points to a less settled browsing rhythm. The keystroke-gap metric adds smaller red pulses on several days including a cluster around early April. We included this keystroke-gap feature in the evaluation to demonstrate the kind of attention-related signal our framework can handle. However, with Dataset B and our indirect extraction path we cannot be certain that the observed values faithfully reflect true inter-keystroke pauses, so we treat its contributions as illustrative rather than definitive until validated on labeled, in-situ data. The smaller red pulses on several days of the keystroke-gap are valuable because they arise from a different behavioral channel than DNS requests and therefore corroborate the network pattern without overlapping it. Where the light-blue retries and the dark-blue burst rate climb together, and a red pulse appears on top, the combined contribution approaches the top of the stack. When only a

single metric is elevated, the combined daily score remains deliberately modest, because the weights and membership curves are tuned to require agreement from several signals before the criterion strengthens. This conservative behavior can reduce false positives.

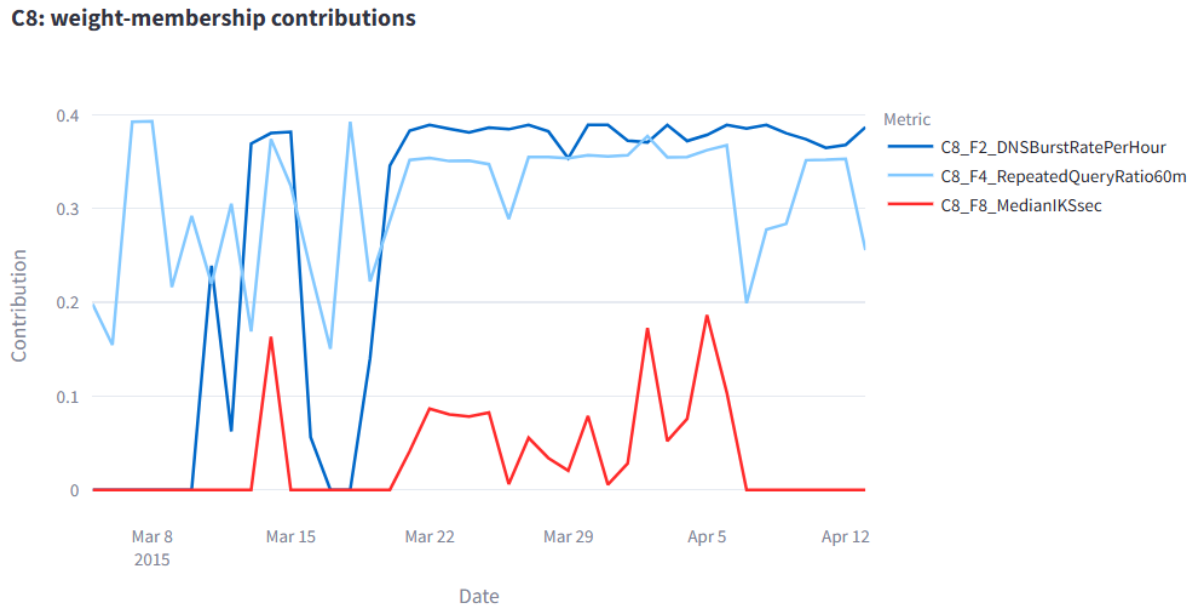


Figure 6.42: C8 weight–membership contributions over time. A repeated-query episode dominates early, a sustained DNS-burst plateau follows, and keystroke gaps add corroborating pulses.

The aggregated likelihood in Figure 6.43 follows this interaction closely. After a gradual rise the curve crosses the decision line and stays mostly in the red band for a multi-week stretch peaking above 0.9 when bursty DNS activity and repeated queries coincide. The brief dips match days where only one metric contributes. This pattern is consistent with the individual membership panels and it is easy to explain to a non-technical user like a parent. The criterion increases when several independent signs of scattered attention align and it relaxes when only one sign fluctuates.

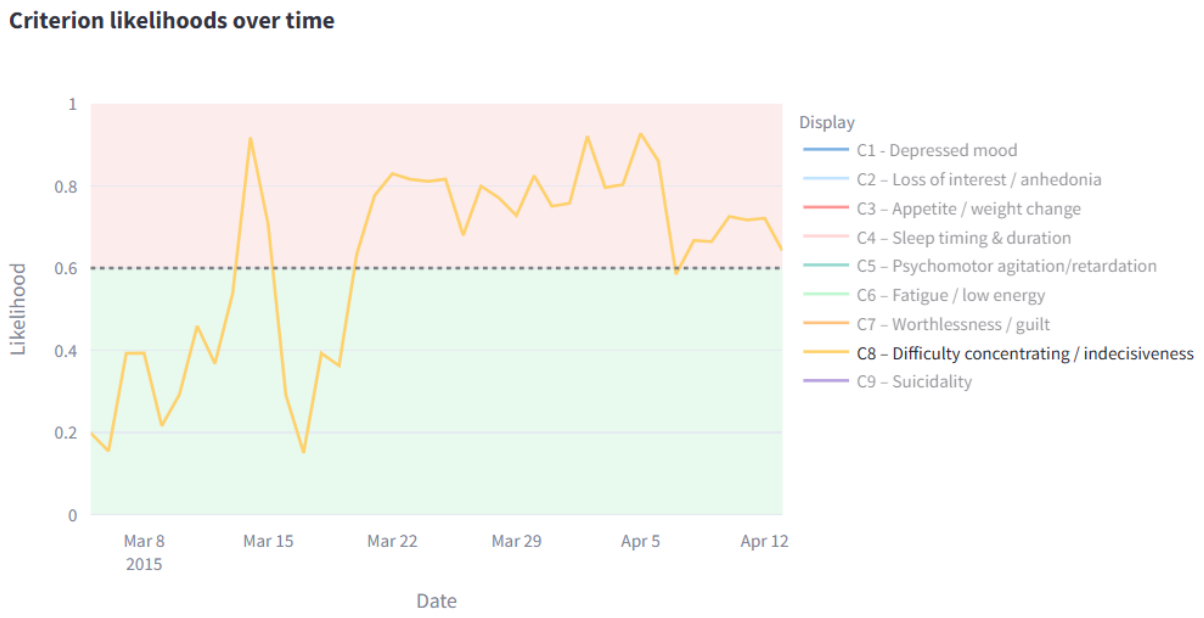


Figure 6.43: C8 criterion likelihood over time. Multi-metric alignment produces sustained periods above the gate; isolated fluctuations only nudge the curve.

The gauge in Figure 6.44 summarizes the 40-day average at 0.622 which is slightly above the criterion gate. This matches the long period with sustained contributions and serves as a clear entry point for a clinician or care-giver. It suggests that attention-related behavior deserves a closer look and the contribution plot immediately shows which aspects dominate on which days. As with the other criteria the same aggregation can later be extended with additional privacy-preserving household indicators that capture focus and decision making from complementary angles.

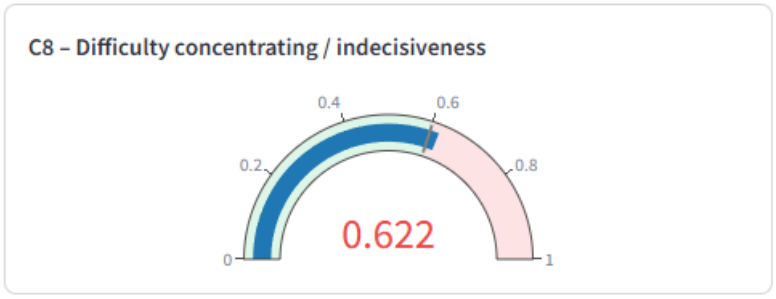


Figure 6.44: C8 40-day average likelihood. The mean is slightly above the gate and consistent with the sustained mid-period elevations.

6.3.5 User Journey

The overview presents the DSM-5 gate in a compact and readable form, and it immediately communicates three decisions: whether an episode is likely, whether a core symptom is satisfied, and how many criteria are currently present within the rolling window. These decisions are computed with the same configuration we used throughout (e.g., $M = 14$, $N = 6$, $\theta = 0.6$, $\tau = 1$) and therefore link transparently to the per-criterion plots above. The design goal is simple. A single glance should tell the user whether there is anything to worry about and where to look next, without hiding how the decision was made.

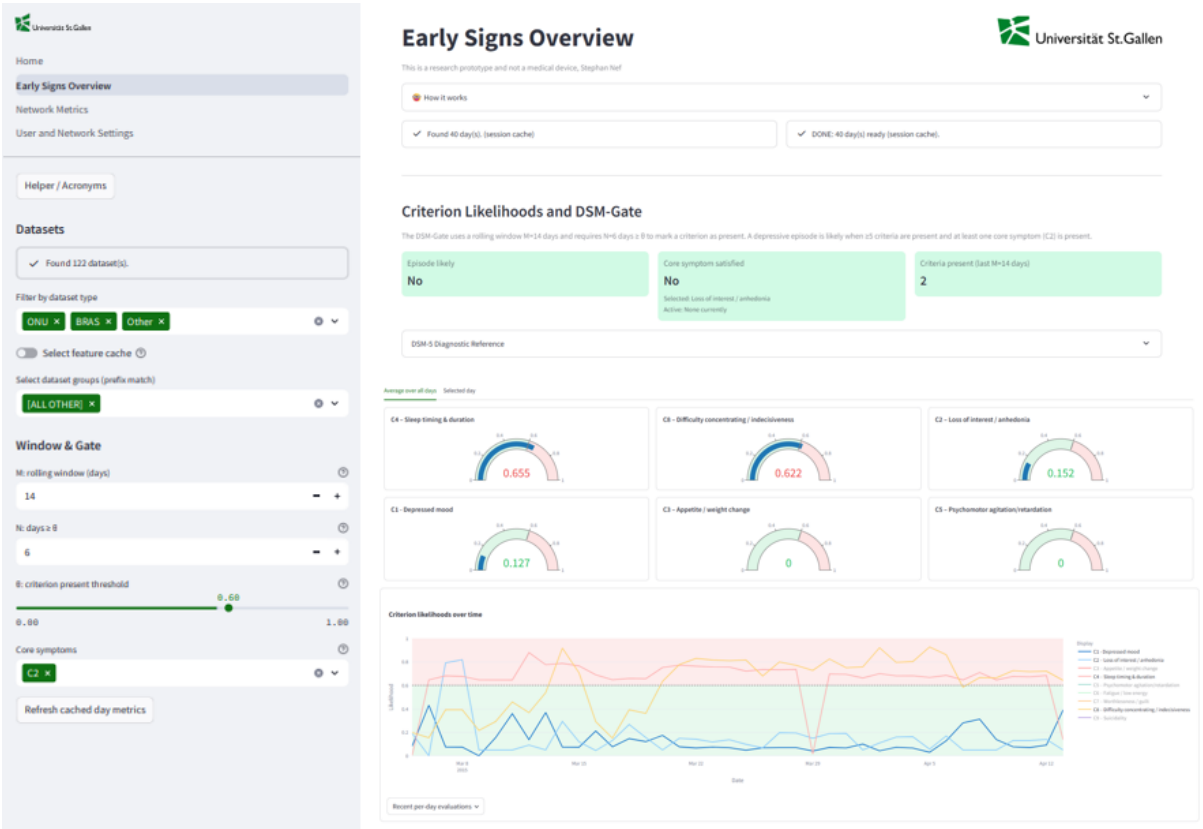


Figure 6.45: Entry Point 'Early Signs Overview'. The page surfaces the DSM-5 gate and places the most important summaries at the top so that triage is effortless.

Figure 6.46 shows the compact episode summary that anchors the page. With the present configuration the application reports that an episode is not likely, no core symptom is satisfied, and two criteria are present in the last fourteen days. This agrees with the detailed sections above. C4 and C8 are elevated for sustained periods while C1 and C2, the two main symptoms, remain below their gates most days. The box clearly explains why the global decision is negative. Our gate logic follows the spirit of DSM-5 and avoids false positives by requiring corroboration and at least one core symptom.

The logic implemented here mirrors the diagnostic thresholds used in DSM-5 for a Major Depressive Episode, where at least five symptoms must be present during the same two-week period, and at least one of these must be a core symptom-either depressed mood (C1) or loss of interest/pleasure (C2). By translating this rule into a transparent, rule-based gate, the system reproduces the DSM-5 structure in computational form. It counts criteria above a defined likelihood threshold ($\theta = 0.6$) within a rolling window ($M = 14$ days) and checks whether any core symptom exceeds the same gate. The decision shown, with two active criteria but no core symptom, would therefore not qualify for a positive episode, in agreement with the textual definition of DSM-5 MDD. This result shows that the framework complies with the specified diagnostic limits while remaining fully explainable and verifiable.

From a clinical perspective, this implementation does not aim to replace professional judgment or diagnosis but to add a continuous, data-driven layer of observation between appointments. We can see within this evaluation that it can provide an additional source of structured evidence that can support existing clinical approaches such as patient interviews, standardized questionnaires (e.g. PHQ-9), or ecological momentary assessments. The system allows for tracking behavioral changes in digital activity that may precede or accompany symptom patterns identified using traditional methods. The transparent link between each metric, its membership function, and the DSM-5 gate ensures that the resulting indicators can be understood and discussed with clinicians, rather than treated as opaque algorithmic outputs. In this way, the approach complements current diagnostic workflows by offering early privacy-preserving signals that align with the DSM-5 MDD categories but remain verifiable through human interpretation and context.
Below the summary, the gauges provide a quick ranking by the average likelihood across the window (Figure 6.47). The viewer sees immediately that the timing and duration of sleep (C4, mean 0.655) and the difficulty of concentrating or indecisiveness (C8, mean 0.622) are the highest and therefore deserve first attention. The loss of interest (C2, mean 0.152) and the depressed mood (C1, mean 0.127) remain low, which matches our per-metric analyzes. This ordering is useful in practice. A clinician can start with the top gauges, drill into the contribution view for those criteria, and only then move on to the lower-ranked ones. Because each gauge is a simple average of daily memberships, and each membership is determined by plots the reader has already seen, the path from the headline number back to raw packet evidence remains transparent.

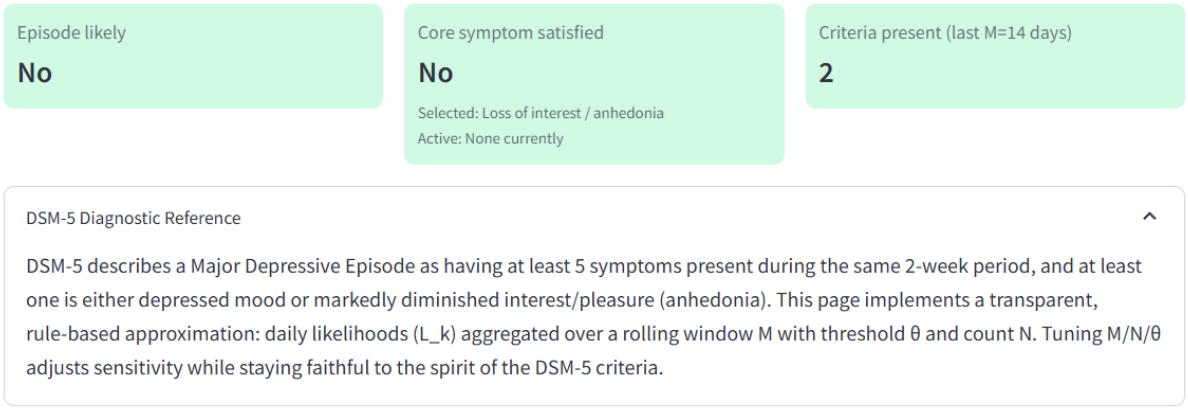


Figure 6.46: Episode Summary. A rule-based approximation of DSM-5 with an explicit reference panel. The decision is immediately visible and parameter choices are shown to support auditability.

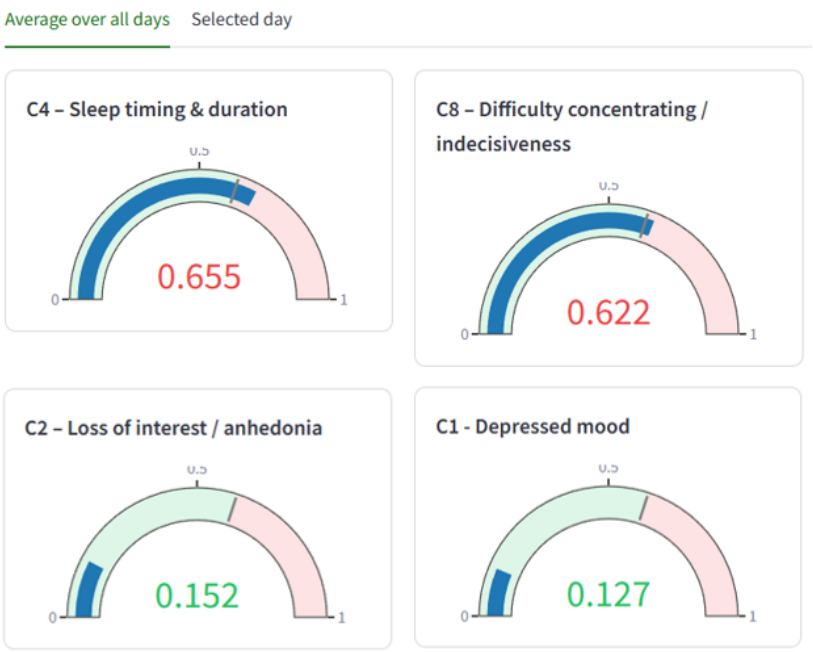


Figure 6.47: Sorted Gauges Overview. A ranked overview that immediately highlights which criteria are persistently elevated (here C4 and C8) and which are quiet (here C1 and C2).

Finally, the time-series of all criteria brings the story together (Figure 6.48). We see that C4 and C8 rise in tandem over a multi-week stretch, that C2 shows a brief early episode and then remains low, and that C1 stays in the green band with only small pulses. This pattern is exactly what the contribution stacks in the criterion sections already suggested.

The overview therefore works as intended. It offers a reliable first read and points to the right places for explanation. If a threshold, weight or limit is changed, the effect is visible here at the top but remains traceable through the same figures that were used to justify the original settings.

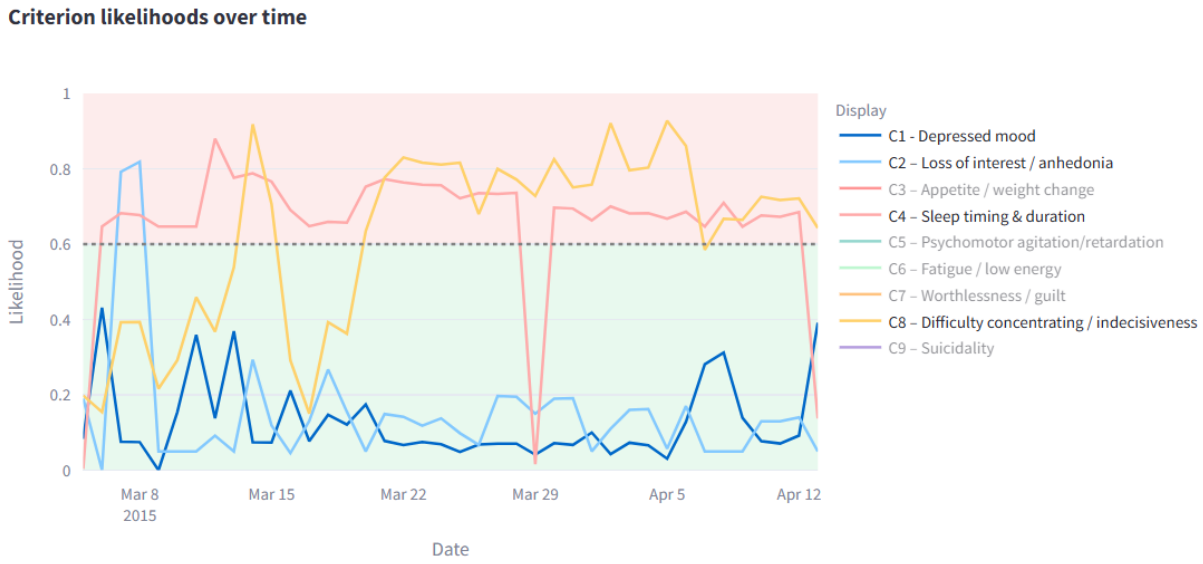


Figure 6.48: All criteria over time. A single plot that aligns the criterion curves so that co-movement and regime changes are easy to spot.

6.4 Discussion

Within this Section, we discuss what the evaluation reveals about our approach. We return to the guiding research question introduced in the beginning of the thesis:

1. Can a router centric pipeline reliably attribute device level and environmental signals to a single individual in a multi person household?
   Yes, but not fully implemented but supported. The prototype does not perform person level allocation. The design supports it through device profiles, stable identifiers, SNI patterns, and time of day regularities. In practice, this provides a workable path for per device attribution and profile mapping. Full person level resolution is future work.

2. Do these signals, when aggregated, correlate with indicators of depressive behavior?
   Yes, we implemented the full router centric pipeline from packet headers to metrics, memberships, and an auditable DSM style gate. We evaluated on two datasets

and showed consistent covariation for sleep timing and duration and for difficulty concentrating. Loss of interest and depressed mood show short and explainable pulses. The gate behaved stably and results are reproducible in the app.

3. Can these indicators enable parents, caregivers, and clinicians to observe early indicators of depression while preserving privacy and meeting ethical safeguards? **Yes**, within the current scope. Processing stays local at the router and only headers and day level features are used. The UI keeps baselines, limits, and contributions transparent for parents, caregivers, and clinicians. This enables early awareness without content inspection.

There is substantial room for future work on clinical validation, person level attribution, and auto tuning. We outline these steps in Chapter 7 and Section 7.2.

With only packet headers, we computed reproducible metrics, mapped them to simple membership functions, and combined them with an auditable gate that mirrors DSM-5 structure. The implemented application with the guiding plots and UI elements made the reasoning traceable from a gauge value back to raw evidence. Where Dataset B constrained certain inferences, the workflow still supported cautious interpretation and graceful down-weighting of affected metrics [36].
The evaluation demonstrates three aspects that matter in practice. First, the method is *explainable by design*. Each metric comes with an explicit baseline (`lo`/`mid`/`hi`), a visible membership curve and a time-series of outputs. Users can see how moving a limit or changing a weight changes the indicator. Second, the approach is *person-centric*. Healthy baselines are set per user and outliers above `hi` are intentionally ignored, which reduces noise while keeping unusual days visible in the raw plots. Third, the DSM-style gate aggregates day-level likelihoods in a way that "feels familiar" to clinicians: it requires corroboration across metrics and at least one core symptom before declaring an episode likely. This is not a diagnostic device. It is an additional, privacy-preserving source of structured observations that can be discussed alongside patient interviews, standardized questionnaires, or ecological momentary assessments.

The evaluation of metrics mapped to the DSM-5 MDD criteria make this concrete. For sleep timing and duration (C4) and difficulty concentrating / indecisiveness (C8) we observed sustained elevations, which the overview surfaced immediately. Depressed

mood (C1) and the loss of interest (C2) remained low except for short, explainable episodes. This separation is useful clinically: it prioritizes where to look first and keeps the path from a high-level decision back to day-level contributions and raw traces unbroken. Where Dataset B produced artifacts (e.g., near-constant wake-after-04:00 due to continuous packet flow), the tool made the issue visible and gave us safe mitigations (down-weighting, thresholding), preserving the overall interpretability.

Beyond the four criteria analyzed in depth, we can say that the workflow scales. The same visual language and the same controls applied to social, typing, diurnal, and DNS-level metrics. This addresses a key gap identified earlier: bridging low-level network signals to clinically meaningful categories with *transparent* logic rather than opaque models. We therefore see the prototype as a practical step toward explainable digital phenotyping at a household vantage.

6.4.1 Privacy Reflection

Both network metadata and mental-health indicators are extremely sensitive information. The edge-first architecture ensures that all computation and data storage stay within the household. Only packet headers and derived, day-level features are used, while payloads are never inspected and no data leave the home by default. This approach minimizes risk, aligns with data-minimization principles, and increases user trust. The strict privacy requirements shaped the design in a positive way. By focusing on content-agnostic features such as timestamps, directions, and SLD hostnames, every step of the analysis became interpretable and easy to audit. This confirms that privacy-by-design does not have to limit insight but rather encourages explainable computation. The trade-off is that less content information requires accurate device attribution and transparent baselines to avoid wrong conclusions, which are both areas for further improvement mentioned in the next Chapter 7.

The prototype already includes several safeguards that ensure privacy and reliability when working with sensitive data such as mental-health-related information:

- Only aggregated, day-level results and curated domain names are stored to avoid exposure of raw data.
- Raw PCAP files can be rotated or deleted automatically after processing to minimize retention of sensitive material.

- Days with low data coverage are flagged and excluded from the DSM-style gating to prevent misleading interpretations.
- All parameter changes are logged to maintain full traceability and allow transparent result verification.

Together, these safeguards ensure that the analysis remains local, transparent, and reversible-an essential requirement for ethical handling of sensitive mental-health-related data.

6.5 Limitations

Our thesis emphasized transparency and operational feasibility from the point of view of the home-router. Several limits qualify the findings and motivate the next steps.

6.5.1 Router-only Observation

Traffic that bypasses the home gateway (mobile data, VPNs, enterprise tunnels) is not visible, which can attenuate or delay indicators. Where feasible, we prefer within-person baselines and we mark days with insufficient evidence in the UI.

6.5.2 Attribution in shared Environments

Household devices may be shared, complicating person-level attribution. The prototype supports profile mapping and device catalogues, but a robust identity-resolution component and systematic handling of shared devices require prospective testing.

6.5.3 Inference without Clinical Ground Truth

Network-derived features only indicate behavior. They do not establish mood state or diagnosis. In this MVP we report criterion-level likelihoods and a DSM-style gate. Controlled studies that collect clinician-rated outcomes in parallel are required to calibrate thresholds and validate utility.

6.5.4 Generality of Datasets and Settings

Results rely on realistic but limited traces curated for research and teaching. The Swedish hands-on dataset provided valuable continuity, yet it does not reflect a single person daily life and contains always-on traffic. Broader ecological validity demands diverse households, longer windows, and cross-site replication.

6.5.5 Model Simplifications

We intentionally used additive fuzzy memberships and an explicit gate to keep the system auditable. We did not include black-box models, and we only used a subset of the BOM logic defined in Chapter 4. Further studies should compare how much accuracy could be gained by using more complex model combinations and how this would affect the system's current level of transparency and interpretability.

6.5.6 Metric-specific Uncertainty

For some signals (e.g., inter-keystroke gaps (C8_F8, B.8) from indirect extraction used in the evaluation for criterion) the numeric reliability in Dataset B is uncertain. We therefore used modest weights and treated contributions as illustrative. Validated and labeled typing data would allow us to tune these metrics more confidently.

All these mentioned limits do not invalidate our approach. They clarify where the system performs well today and where further development is required before real-world deployment. The open issues identified here correspond to the remaining parts of the research gap discussed in the Introduction and will be addressed in the next Chapter (7), specifically in Section 7.2.

Chapter 7

Conclusion & Future Work

The preceding evaluation and discussion have shown that a router-centric, privacy-first approach can transform everyday network data into interpretable behavioral indicators aligned with DSM-5 MDD criteria. Having examined the systems feasibility, transparency and current boundaries, this final chapter steps back to consolidate what has been achieved and outline where the approach can grow. It summarizes the main insights from implementation and evaluation, sum up the core contributions, and identifies the next technical, methodological, and ethical steps needed to advance from a MVP toward a clinically meaningful and deployable system. The following sections therefore present, first, the main conclusions and key learnings and, second, a roadmap for future research and development.

7.1 Conclusion

This thesis does not claim a clinical breakthrough, and it also does not replace existing diagnostic practice. However, it delivers a solid, working basis on which such work can grow. Starting from the research question, we designed and implemented a pipeline that turns packet headers into day-level, criterion-aligned likelihoods, with adjustable, explainable parameters and an auditable DSM-style gate. The evaluation showed that the system surfaces plausible episodes, that outliers can be handled transparently, and that the overview and contribution plots give non-technical users a clear entry point. Where the data constrained us, we remained explicit about assumptions and mitigations.

In terms of research gaps, the work advances three out of four that were identified in Chapter 1, Section 1.3. First, it closes the *methodological bridge* gap by providing an auditable mapping from home network signals to DSM-5–aligned day-level indicators (Chapter 4, Section 6.2). Second, it addresses the *explainability and personalization* gap by showing how person-specific baselines and limits (`lo`/`mid`/`hi`) can be tuned without

opaque models, with every choice visible in the membership and contribution plots. Third, it tackles the *operationalization and reproducibility* gap with a privacy-preserving UI and stored configurations that let clinicians and care-givers retrace results and rerun analyzes on demand. But also at the same time, the absence of labeled, in-situ ground truth limits external validity and forced conservative assumptions for some metrics, which likely dampened numerical results.

Taken together, these contributions answer the research question in the affirmative. A router-centric, packet-header–only system can produce explainable, DSM-5 MDD consistent indicators that complement clinical practice. The prototype is not a diagnostic instrument and does not replace professional judgment. It is a solid working basis on which to build. With labeled field data, stronger identity resolution, and explainable auto-tuning, the same design can be extended to quantify sensitivity/specificity against clinician-rated outcomes and to assess the added value of digital-environment signals alongside traditional assessments. The implementation also revealed how fragile person-level attribution can be in shared environments. Even small errors in IP-to-user mapping propagate through the logic, emphasizing that identity resolution is not a technical add-on but the foundation for any credible behavioral inference

7.2 Future Work

The evaluation surfaces several concrete opportunities.

7.2.1 Prospective Field Studies with Ground Truth

The most important next step is a longitudinal field study that collects (i) router-level traces, (ii) person-level labels via standardized instruments (e.g., PHQ-9) and clinician interviews, and (iii) brief participant context (sleep schedules, travel, illnesses). This would provide a defensible ground truth both on the clinical and digital side - and would allow proper calibration of thresholds, weights, and limits. It would also allow sensitivity/specificity analysis of the DSM-style gate and criterion aggregation.

7.2.2 Auto-tuning of Parameters

Complement manual, profile-aware tuning with automatic but interpretable methods. Candidates include clustering, such as k-means [30, 29] or Gaussian Mixture Models

fitted via the Expectation–Maximization algorithm [15, 7], to propose person-specific baselines from recurring behavioral patterns. Density-based clustering methods like DBSCAN [17] can identify stable regimes and highlight outlier days that differ markedly from the individual's routine. Seasonal trend decomposition (STL) [13] or exponentially weighted smoothing approaches [43, 22] could dynamically adjust `lo`/`mid`/`hi` to account for weekday or seasonal effects. Change-point detection algorithms such as Bayesian online change-point detection [1] or CUSUM [38, 54] can update thresholds automatically when the user's behavior changes over time. Finally, monotonic or shape-constrained Generalized Additive Models [40, 62] could be used to learn smooth, human-readable membership curves while maintaining interpretability. We implemented a first prototype using a simple clustering-based baseline detector, but the results in Dataset B were not yet convincing, most likely due to limited and continuous input data, and therefore further refinement was postponed due to time constraints. Therefore, we propose to continue this work on labeled field data collected under real-world conditions, where these explainable data-mining approaches can be validated and integrated more effectively.

7.2.3 Metric Catalog and Behavior Observation Metrics

In future work, we can scale the metric catalog beyond router traffic by integrating additional digital environmental signals while keeping the design explainable and privacy-first. A possible promising next step could be to merge our router-centric indicators with an audio-centered signal stream as proposed by Länzlinger [28]. In our framework, such audio features would leverage our existing approach and preserve the same transparent logic.

Beyond audio, further signals can be included where consent and privacy safeguards allow, as example, but not limited to:

- **Lights and luminance** (on/off events, light level), to refine sleep timing and day–night rhythm.
- **Temperature and humidity**, to provide context for sleep comfort and routine stability.
- **Motion or room presence** sensors, to strengthen evidence for activity, inactivity, or nocturnal awakenings.

Any added modality should pass a privacy re-evaluation before inclusion.

7.2.4 Identity Resolution

Extend device–profile mapping and support shared devices, then, where consent allows, we could combine home-router traces with privacy-preserving summaries from audio data, phones, or wearables (e.g. daily active hours or coarse activity phases) to disambiguate metrics while keeping payloads private [28, 56, 44, 21]. Even lightweight cross-signals can clarify who was active when, reduce ambiguity from shared devices, and strengthen the criterion evidence without compromising the content-agnostic design.

7.2.5 Robustness, Fairness, and Governance

These safeguards make the approach reliable enough for real households and reliable and trustworthy enough for clinical use. For robustness, we will add data-quality tracking (e.g. per-day percentage coverage, VPN/tunnel detection, outage markers), conservative fallbacks (e.g. widen M, suppress gating on low-coverage days) and simple, explainable imputations (carry-forward within capped windows) with visible confidence flags in the UI. This reduces false positives from missing data and stabilizes longitudinal trends, directly improving the usefulness of criterion likelihoods. For fairness, we could test metrics across different profile settings (household size, device mix, ISP/router class, home-office patterns) and report splitted performance so weights and limits can be adjusted where systematic bias appears. This ensures that a threshold learned on one household type does not over- or under-signal in another, increasing generalizability. For governance, we will formalize consent, data minimization, and retention policies (e.g., store only eTLD+1 and counts, rotate raw PCAP files, provide audit logs of parameter changes, and enable subject access/erasure). Remaining content-agnostic by design strengthens privacy and regulatory alignment, which in turn increases clinical acceptability and adoption. Together, these measures raise confidence in the indicators, reduce operational risk, and allow our explainable workflow to be deployed and compared fairly in diverse settings.

Overall, this thesis offers a practical and transparent foundation for explainable early-sign indicators from a home-router perspective. We showed that packet-header signals can be transformed into DSM-5 MDD aligned Day-level likelihoods that remain auditable from high-level indicators to packet trace. The evaluation shows promise, the limitations are clear, and the path to higher validity is well defined. With real-world labeled studies, careful auto-tuning, an expanded catalog of privacy-preserving metrics, and close collaboration with clinicians and families, this approach can mature into a robust companion to

established clinical practice. It adds a continuous, comprehensible context while neither obscuring nor replacing a professional clinical diagnosis.

7.2.6 Lessons Learned

Throughout the development of this system design and the implementation of the MVP, several key lessons emerged that go beyond the immediate technical achievements.
First, the work demonstrated that explainability and privacy are not opposing goals but mutually reinforcing design principles. By limiting ourselves to packet headers and visible metadata, we were forced to make every transformation explicit and interpretable. This constraint led to a cleaner and more reliable architecture, confirming that meaningful behavioral insight can be achieved without compromising user privacy.

Second, the project revealed how critical identity resolution is for any form of user-centric behavioral analysis. Even minimal uncertainty in device-to-user mapping propagated through the pipeline and influenced downstream interpretations. This experience taught us that technical precision in attribution is as essential as methodological rigor in the psychological mapping of features to DSM-5 MDD criteria.

Third, the process of mapping low-level network metrics to Behavior Observation Metrics (BOM) and further to DSM-5 MDD criteria proved to be one of the most instructive aspects of the project. It revealed how much conceptual discipline is required to bridge digital behavior and clinical constructs without overfitting or speculation. Each mapping step demanded clear justification, consistent terminology, and transparency in assumptions. This exercise underscored that diagnostic interpretability emerges not from algorithmic sophistication but from a well-documented reasoning chain connecting measurable behavior to recognized symptom categories.

Fourth, we learned that simplicity often outperforms complexity when transparency is a priority. The additive fuzzy memberships and DSM-style gating mechanisms may appear basic compared to modern machine-learning models, yet their interpretability made collaboration with non-technical stakeholders possible. Clinicians, for example, could immediately understand how each metric contributed to a likelihood at the criterion-level.

Finally, we realized that building such a system requires a shift in mindset from maximizing data collection to maximizing understanding per data bit. Every architectural decision, from membership functions to day-level aggregation, reinforced the idea that less can be more when the transformation chain remains auditable.

These lessons form the conceptual foundation for future extensions and are likely to be transferable to other privacy-critical domains where behavioral inference meets ethical responsibility.

Bibliography

[1] R. P. Adams and D. J. C. MacKay. “Bayesian Online Changepoint Detection”. In: *arXiv preprint arXiv:0710.3742* (2007).

[2] Doaa Alamoudi, Ian Nabney, and Esther Crawley. “Evaluating the effectiveness of the SleepTracker app for detecting anxiety-and depression-related sleep disturbances”. In: *Sensors* 24.3 (2024), p. 722.

[3] Tim Althoff et al. “Large-scale diet tracking data reveal disparate associations between food environment and diet”. In: *Nature communications* 13.1 (2022), p. 267.

[4] Abayomi Arowosegbe and Tope Oyelade. “Application of natural language processing (NLP) in detecting and preventing suicide ideation: a systematic review”. In: *International Journal of Environmental Research and Public Health* 20.2 (2023), p. 1514.

[5] Pradeep K Atrey et al. “Multimodal fusion for multimedia analysis: a survey”. In: *Multimedia systems* 16.6 (2010), pp. 345–379.

[6] Philippe Biondi and contributors. *Scapy—Packet Manipulation Tool*. 2025.

[7] C. M. Bishop. *Pattern Recognition and Machine Learning*. 2006.

[8] Ana-Maria Bucur. “Leveraging LLM-generated data for detecting depression symptoms on social media”. In: *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer. 2024, pp. 193–204.

[9] Carter Bullard. *Argus: Audit Record Generation and Utilization System*. https://qosient.com/argus/. Network flow monitoring tool for real-time and historical traffic analysis. 2024.

[10] Lora E Burke, Jing Wang, and Mary Ann Sevick. “Self-monitoring in weight loss: a systematic review of the literature”. In: *Journal of the American Dietetic Association* 111.1 (2011), pp. 92–102.

[11] Fidel Cacheda et al. “Early detection of depression: social network analysis and random forest techniques”. In: *Journal of medical Internet research* 21.6 (2019), e12554.

[12] Zahra M Clayborne, Melanie Varin, and Ian Colman. “Systematic review and meta-analysis: adolescent depression and long-term psychosocial outcomes”. In: *Journal of the American Academy of Child & Adolescent Psychiatry* 58.1 (2019), pp. 72–79.

[13] R. B. Cleveland et al. “STL: A Seasonal–Trend Decomposition Procedure Based on Loess”. In: *Journal of Official Statistics* 6.1 (1990), pp. 3–73.

[14] Munmun De Choudhury et al. “Discovering shifts to suicidal ideation from mental health content in social media”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2016, pp. 2098–2110.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: *Journal of the Royal Statistical Society: Series B* 39.1 (1977), pp. 1–38.

[16] Docker Inc. *Docker*. 2025.

[17] M. Ester et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In: *Proc. KDD*. 1996, pp. 226–231.

[18] *Federal Act on Data Protection (FADP), SR 235.1*. FedLex. 2025. URL: https://www.fedlex.admin.ch/eli/cc/2022/491/en.

[19] Lauren Hale and Stanford Guan. “Screen time and sleep among school-aged children and adolescents: a systematic literature review”. In: *Sleep medicine reviews* 21 (2015), pp. 50–58.

[20] Erik Hjelmvik and Swedish Armed Forces CERT. *Hands-on Network Forensics*. Workshop material. FIRST 2015 Berlin, including 40-day continuous capture dataset. 2015. URL: https://share.netresec.com/s/NBbonFnsZ4HqCTW?dir=/&editing=false&openfile=true.

[21] J. F. Huckins et al. “Fusing Mobile Phone Sensing and Ecological Momentary Assessments to Assess Depression in College Students”. In: *Science Advances* 6.36 (2020), eaba1944.

[22] R. J. Hyndman et al. *Forecasting with Exponential Smoothing: The State Space Approach*. 2008.

[23] IAB Tech Lab. *IAB Tech Lab Content Taxonomy v2.2*. https://iabtechlab.com/standards/iab-content-taxonomy/. 2023.

[24] *ISO/IEC 31700-1:2023 — Consumer protection: Privacy by design for consumer goods and services, Part 1: High-level requirements*. International Organization for Standardization and International Electrotechnical Commission, 2025.

[25] Wei Jiang et al. “A real network environment dataset for traffic analysis”. In: *Scientific Data* 12.1 (2025), p. 756.

[26] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. “The PHQ-9: validity of a brief depression severity measure”. In: *Journal of general internal medicine* 16.9 (2001), pp. 606–613.

[27] Kostadin Kushlev, Jason Proulx, and Elizabeth W Dunn. “"Silence your phones" Smartphone notifications increase inattention and hyperactivity symptoms”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2016, pp. 1011–1020.

[28] Jonas Laenzlinger. “Audio-centered Approach for Building a Multimodal Predictive AI Agent to Detect Depressive Behaviors”. Master Thesis. University of St. Gallen, School of Computer Science, 2025.

[29] S. P. Lloyd. “Least Squares Quantization in PCM”. In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 129–137.

[30] J. MacQueen. “Some Methods for Classification and Analysis of Multivariate Observations”. In: *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. 1967, pp. 281–297.

[31] Priyanka Mary Mammen et al. “Wisleep: Inferring sleep duration at scale using passive wifi sensing”. In: *arXiv preprint arXiv:2102.03690* (2021).

[32] Rafail-Evangelos Mastoras et al. “Touchscreen typing pattern analysis for remote detection of the depressive tendency”. In: *Scientific reports* 9.1 (2019), p. 13414.

[33] Patrick D McGorry and Cristina Mei. “Early intervention in youth mental health: progress and future directions”. In: *BMJ Ment Health* 21.4 (2018), pp. 182–184.

[34] Mozilla Foundation. *Public Suffix List: Documentation*. https://publicsuffix.org/. 2024.

[35] Katharina O. E. Müller et al. “Big Brother is Watching You: Non-Intrusive ZigBee User Profiling”. In: *Proceedings of the 20th International Conference on Network and Service Management (CNSM)*. 2024.

[36] Stephan Nef. *app_pcap_to_dsm5_public: PCAP → DSM-5 Streamlit prototype*. https://github.com/nefste/app_pcap_to_dsm5_public. master thesis codebase; includes implemented app, metrics, minimal ETL process, feature collection and Docker/Compose assets. 2025.

[37] Yoshiyo Oguchi et al. “Potential Predictors of Delay in Initial Treatment Contact After the First Onset of Depression in Japan: A Clinical Sample Study”. In: *International Journal of Mental Health Systems* 8 (2014), p. 50. DOI: 10.1186/1752-4458-8-50.

[38] E. S. Page. “Continuous Inspection Schemes”. In: *Biometrika* 41.1/2 (1954), pp. 100–115.

[39] Steffen H Pedersen et al. “ICD-10 criteria for depression in general practice”. In: *Journal of affective disorders* 65.2 (2001), pp. 191–194.

[40] N. Pya and S. N. Wood. “Shape Constrained Additive Models”. In: *Statistics and Computing* 25 (2015), pp. 543–559.

[41] Python Software Foundation. *ipaddress — IPv4/IPv6 manipulation library*. Python Software Foundation. 2025.

[42] *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union. 2025. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ%3AL_202401689.

[43] S. W. Roberts. “Control Chart Tests Based on Geometric Moving Averages”. In: *Technometrics* 1.3 (1959), pp. 239–250.

[44] S. Saeb et al. “The Relationship Between Mobile Phone Location Sensor Data and Depressive Symptom Severity”. In: *PLOS ONE* 10.7 (2015), e0127468.

[45] Sohrab Saeb et al. “Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study”. In: *Journal of medical Internet research* 17.7 (2015), e4273.

[46] Sohrab Saeb et al. “Scalable passive sleep monitoring using mobile phones: opportunities and obstacles”. In: *Journal of medical Internet research* 19.4 (2017), e118.

[47] Johanne Smith-Nielsen et al. “Validation of the Edinburgh Postnatal Depression Scale against both DSM-5 and ICD-10 diagnostic criteria for depression”. In: *BMC psychiatry* 18 (2018), pp. 1–12.

[48] *Snort: Open Source Intrusion Prevention System (IPS)*. https://www.snort.org/. Snort is a free and open-source network intrusion detection and prevention system (IDS/IPS) capable of performing real-time traffic analysis and packet logging. 2025.

[49] Clemens Stachl et al. “Predicting personality from patterns of behavior collected with smartphones”. In: *Proceedings of the National Academy of Sciences* 117.30 (2020), pp. 17680–17687.

[50] Streamlit Inc. *Manage Your App — Streamlit Community Cloud*. Accessed: 2025-10-18. 2025. URL: https://docs.streamlit.io/deploy/streamlit-community-cloud/manage-your-app.

[51] Streamlit Inc. *Streamlit*. 2025.

[52] Hajime Sueki. “The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan”. In: *Journal of affective disorders* 170 (2015), pp. 155–160.

[53] Fredrik Svenaeus. “Diagnosing mental disorders and saving the normal: American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders, American Psychiatric Publishing: Washington, DC. 991 pp., ISBN: 978-0890425558. Price: 122.70”. In: *Medicine, Health Care and Philosophy* 17 (2014), pp. 241–244.

[54] C. Truong, L. Oudre, and N. Vayatis. “Selective Review of Offline Change Point Detection Methods”. In: *Signal Processing* 167 (2020), p. 107299.

[55] Rudolf Uher et al. “Major depressive disorder in DSM-5: Implications for clinical practice and research of changes from DSM-IV”. In: *Depression and anxiety* 31.6 (2014), pp. 459–471.

[56] R. Wang et al. “StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones”. In: *Proc. UbiComp*. 2014, pp. 3–14.

[57] Rui Wang et al. “StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones”. In: *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 2014, pp. 3–14.

[58] Rui Wang et al. “Tracking depression dynamics in college students using mobile phone and wearable sensing”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.1 (2018), pp. 1–26.

[59] Shweta Ware et al. “Large-scale automatic depression screening using meta-data from wifi infrastructure”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.4 (2018), pp. 1–27.

[60] Shweta Ware et al. “Predicting depressive symptoms using smartphone data”. In: *Smart Health* 15 (2020), p. 100093.

[61] Wireshark Foundation. *Wireshark Network Protocol Analyzer*. 2025.

[62] S. N. Wood. *Generalized Additive Models: An Introduction with R (2nd ed.)* 2017.

[63] *Zeek: A Free and Open Source Network Analysis Framework*. https://zeek.org/. Zeek is a free and open-source software network analysis framework used for network monitoring and security analysis. 2025.

[64] John Zulueta et al. “Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffect digital phenotyping study”. In: *Journal of medical Internet research* 20.7 (2018), e241.

Appendix A
Overview Related Work

Table A.1: Overview of related work structured by the strands used in Chapter 3. Gaps (G*n*) and Contributions (C*n*) addressed by the thesis are highlighted in Chapter 3. The rightmost column provides per-work notes referenced from Table 3.1.

Reference	Contribution	Relevance & Notes (for Table 3.1)
1. Passive Digital Sensing		
"StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones" [57]	Smartphone app linking mobility, sleep, and social rhythm to weekly mood in 48 students.	Device-centric baseline for passive monitoring; motivates router alternative for household settings (G2). One-device-per-user assumption; weekly aggregation; no router vantage; no DSM-style decision; privacy depends on app permissions. **1**
"Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study" [45]	GPS/accelerometer features explain a notable share of PHQ-9 variance.	Predictive value from phones; lacks multi-user attribution and clinical mapping (G2, G4). Single-user capture; feature–PHQ correlation without criterion mapping; no network metadata; no privacy-by-design guarantees. **2**

Continued on next page

(continued from previous page)

Reference	Contribution	Relevance & Notes (for Table 3.1)
"Tracking depression dynamics in college students using mobile phone and wearable sensing" [58]	Phone + Fitbit fusion improves depression detection (N=221).	Fusion inspires multi-source design; still per-user devices (G2). Not router-centric; mixes device/wearable streams; partial interpretability via hand-crafted features; no DSM gate. **3**
"Scalable passive sleep monitoring using mobile phones: opportunities and obstacles" [46]	Six-month wearable study forecasting mood shifts.	Shows longitudinal feasibility; compliance challenges in families. Per-person wearable; no household disambiguation; no payload constraints; no DSM-style logic. **4**
""Silence your phones" Smartphone notifications increase inattention and hyperactivity symptoms" [27]	Notification burst manipulation affects short-term affect.	Inspires burst-based temporal features. Device-level events; partial interpretability (manipulation check) but not clinical mapping; no router vantage. **5**
"Evaluating the effectiveness of the SleepTracker app for detecting anxiety-and depression-related sleep disturbances" [2]	Sleep-tracker metrics predict PHQ-9 one week ahead.	Lead-time on sleep disturbance; motivates metadata-only sleep proxy. App/wearable-based; focuses on sleep only; partial interpretability; no multi-resident attribution. **6**

Continued on next page

(continued from previous page)

Reference	Contribution	Relevance & Notes (for Table 3.1)
"Early detection of depression: social network analysis and random forest techniques" [11]	"Early risk" classifier using 1.8M tweets (AUC 0.82).	Motivates privacy-preserving alternative to textual content. Content analysis; not metadata-only; model not mapped to DSM criteria; privacy trade-offs. **7**
"Discovering shifts to suicidal ideation from mental health content in social media" [14]	Temporal linguistic shifts on Reddit before self-reported episodes.	Time-series framing relevant; payload infeasible here. Content-based signals; partial interpretability via linguistic categories; no DSM gate; not household-aware. **8**
"Leveraging LLM-generated data for detecting depression symptoms on social media" [8]	BERT-based depression detection on Reddit (F1 $\approx$ 0.86).	Shows performance from content; strengthens metadata-only stance. Deep NLP on payload; opaque predictions; no DSM mapping; privacy risk. **9**
"Predicting personality from patterns of behavior collected with smartphones" [49]	Large-scale prediction from smartphone logs and digital records.	Demonstrates breadth of device metadata; lacks network vantage and DSM mapping. Device-level metadata (apps, usage); single-user assumption; partial interpretability; no router vantage; no DSM gate. **10**

Continued on next page

(continued from previous page)

Reference	Contribution	Relevance & Notes (for Table 3.1)
2. Environmental Sensing & Network-Traffic Metadata		
"Large-scale automatic depression screening using meta-data from wifi infrastructure" [59]	Campus WiFi logs: higher PHQ-9 → fewer AP connections and less night mobility.	Proof-of-concept for network metadata; lacks feature-to-symptom mapping (G4). Infrastructure vantage (campus), not home router; metadata-only; partial interpretability; no DSM gate; privacy depends on policy. **11**
"Wisleep: Inferring sleep duration at scale using passive wifi sensing" [31]	Home-gateway traffic gaps infer sleep duration and awakenings.	Validates router sleep proxy; daytime behavior not covered. Router-centric; metadata-only; per-device proxy for one user; partial interpretability (sleep only); no DSM gate. **12**
"Predicting depressive symptoms using smartphone data" [60]	Evening traffic dip predicts PHQ-9 rise one week later.	Adds lead-time; lacks household disambiguation. Home routers; metadata-only; aggregate household signal (no person attribution); partial interpretability; no DSM gate. **13**
Public Suffix List: Documentation [34]	Domain hierarchy from hostnames for privacy-safe post-processing.	Used for domain enrichment; supports metadata-only approach.

Continued on next page

(continued from previous page)

Reference	Contribution	Relevance & Notes (for Table 3.1)
IAB Tech Lab Content Taxonomy v2.2 [23]	Coarse content taxonomy used for high-level content categories.	Used to derive privacy-safe content labels for flow bursts.
3. Clinical Symptom Frameworks (also see Chapter 2)		
"ICD-10 criteria for depression in general practice" [39]	ICD-10 symptom-count thresholds; 50% of episodes missed at first contact.	Motivates pre-clinical resolution (G1); no sensing pipeline.
"Validation of the Edinburgh Postnatal Depression Scale against both DSM-5 and ICD-10 diagnostic criteria for depression" [47]	EPDS psychometrics; recall bias over a two-week window.	Supports the need for continuous, objective monitoring.
"Diagnosing mental disorders and saving the normal: American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders, American Psychiatric Publishing: Washington, DC. 991 pp., ISBN: 978-0890425558. Price: 122.70″	Critique that DSM/ICD capture only snapshots.	Justifies hour-level observability.

Appendix B
Feature Mapping to Behavior Observation Metric

Table B.1: Criterion 1 Feature Mapping – DSM-5 MDD ("Depressed mood most of the day, nearly every day").

Behavior Observation Metric (BOM)	Network-Traffic Feature	Feasible w/ header & metadata	Detail Ref.
(BOM1) Reduced Social Interaction	F1 Distinct social domains	Yes	D1
	F2 Mean social-session duration	Yes	D2
(BOM2) Passive Media Binge	F3 Long streaming flows	Yes	D3
	F4 Down/Up byte ratio	Yes	D4
(BOM3) Rumination Browsing loops	F5 Revisit-ratio	Yes	D5
	F6 Short-interval repeats	Yes	D6
(BOM4) Flattened diurnal Rhythm	F7 Hourly CV	Yes	D7
	F8 Night/Day traffic ratio	Yes	D8

Details for Table B.1

D1. Distinct *eTLD+1* domains / day: Counts how many different sites a device touches in a day; fewer sites = less exploration [58]. Notes / How to Compute: per device/day count unique eTLD+1 from DNS/SNI; report z vs 28-day baseline (fallback to IP/ASN if DoH/ECH).

D2. Mean social-session duration: Average time spent per visit on social apps/sites; longer sessions can reflect lingering, shorter can reflect withdrawal [58]. *Notes / How to Compute:* label social domains; sessionize with 15–30min idle gap; take daily mean.

D3. Long streaming flows ($\geq$2h): Flags extended, passive viewing sessions [11]. *Notes / How to Compute:* label streaming services (hostname/IP/ASN); merge flows with 5min idle gap; mark sessions $\geq$7200s.

D4. Down/Up byte ratio: Passive watching has lots of download and little upload [11]. *Notes / How to Compute:* for streaming sessions compute $R = B_{\downarrow}/B_{\uparrow}$; summarize median / P75 per day.

D5. Revisit-ratio: Repeatedly opening the same few sites suggests looping/rumination [14].*Notes / How to Compute:* in rolling 30min windows, unique eTLD+1 ÷ total DNS/SNI touches; aggregate daily.

D6. Short-interval repeats: Frequent quick returns to the same domain show checking behavior [14]. *Notes / How to Compute:* count repeats $<$ 30min for each domain; compute daily Fano factor $F = \mathrm{Var}(N)/\mathbb{E}[N]$.

D7. Hourly CV (rhythm): Measures day–night pattern strength; lower CV = flatter rhythm [58]. *Notes / How to Compute:* per hour sum active minutes (preferred) or bytes; compute CV across 24 hours; 7–14day smoothing.

D8. Night/Day traffic ratio: More night than day activity indicates nocturnal drift/sleep intrusion [58]. *Notes / How to Compute:* ratio of activity in 22:00–06:00 vs 06:00–22:00; compare to 14-day baseline (use active minutes to reduce streaming bias).

Table B.2: Criterion 2 Feature Mapping – DSM-5 MDD ("Markedly diminished interest or pleasure in almost all activities").

Behavior Observation Metric (BOM)	Network-Traffic Feature (ID)	Feasible w/ header & metadata	Detail Ref.
(BOM1) Shrinking Domain Diversity	F1 Unique eTLD+1 domains / day	Yes	D1
	F2 Domain-category entropy	Yes	D2
(BOM2) Reduced Interactive Engagement	F3 Chat-session count	Partial	D3
	F4 Median reply latency	Partial	D4
	F5 Upstream byte-rate	Yes	D5
(BOM3) Drop in Goal-Oriented Browsing	F6 Productivity-site hits	Yes	D6
	F7 Passive/Active traffic ratio	Partial	D7

Details for Table B.2

D1. Unique eTLD+1 Domains: Daily count of different sites visited; fewer sites over time suggests reduced exploration [34, 58]. *Notes / How to Compute:* per device/day, count unique eTLD+1 from DNS/SNI; report z vs 28-day baseline.

D2. Domain-category Entropy: Variety of topics visited; lower entropy means narrower interests [23, 58]. *Notes / How to Compute:* map eTLD+1 to IAB categories; compute daily Shannon entropy $H = -\sum p_i \log_2 p_i$.

D3. Chat-session Count: Number of messaging sessions; fewer sessions can indicate social withdrawal [58]. *Notes / How to Compute:* label chat domains (WhatsApp/Discord/Slack); sessionize with 5 min idle gap; count per day.

D4. Median Reply Latency: Typical time to respond in chats; slower responses suggest lower engagement [58]. *Notes / How to Compute:* within chat sessions, measure gaps between small upstream and downstream bursts; take median per day.

D5. Upstream Byte-rate: How much a person *sends* during interactive use; sustained drops imply more "lurking" [11]. *Notes / How to Compute:* upstream bytes ÷ session duration for chat sessions; compare to personal baseline.

D6. Productivity-site Hits: Visits to tools like Docs/Notion/Learning Management Systems (LMS)/calendars; fewer visits reflect less goal-directed activity [58]. *Notes / How to Compute:* maintain a productivity domain list; count sessions per day via DNS/SNI/IP labels.

D7. Passive / Active Traffic Ratio: Balance of passive (down-heavy) vs interactive (up-bursty) use; rising ratio signals more passive consumption [11]. *Notes / How to Compute:* define passive = streaming/CDN sessions; active = small frequent upstream bursts in interactive apps; compute daily ratio vs personal percentile.

Table B.3: Criterion 3 Feature Mapping – DSM-5 MDD ("Significant weight loss or gain, or change in appetite nearly every day").

Behavior Observation Metric (BOM)	Network-Traffic Feature (ID)	Feasible w/ header & metadata	Detail Ref.
(BOM1) Food-ordering Pattern Shift	F1 Food-delivery domain hits / week	Yes	D1
	F2 Late-night (22–06h) delivery ratio	Yes	D2
	F3 Mean inter-order interval (days)	Partial	D3
(BOM2) Calorie / Nutrition Information Seeking	F4 Diet-site visit count	Yes	D4
	F5 Calorie-tracker sync bursts	Partial	D5
(BOM3) Smart-scale usage Variability	F6 Smart-scale upload events	Partial	D6
	F7 Weigh-in time variability	Partial	D7

Details for Table B.3

D1. Food-delivery domain hits: Weekly visits to takeaway services; big shifts up/down can reflect appetite change. *Notes / How to Compute:* count sessions to curated delivery eTLD+1; report weekly z vs 4-week baseline.

D2. Late-night delivery ratio: Share of delivery sessions happening at night; higher night eating is linked to atypical patterns [3]. *Notes / How to Compute:* ratio = sessions in 22:00–06:00 ÷ all delivery sessions (rolling weekly).

D3. Mean inter-order interval: Typical gap between delivery *visits*; shorter gaps suggest increased reliance on outside food. *Notes / How to Compute:* sort delivery sessions by time; average days between sessions (domain-level proxy for orders).

D4. Diet-site visit count: Interest in nutrition/fitness content may rise around appetite/weight changes [23]. *Notes / How to Compute:* map eTLD+1 to IAB "Health/Fitness/Nutrition"; count sessions per day or week.

D5. Calorie-tracker sync bursts: More frequent logging/syncing indicates active self-monitoring. *Notes / How to Compute:* detect short sessions to tracker domains with noticeable upstream spikes; count bursts per day (no payload/paths).

D6. Smart-scale upload events: Scale readings often trigger small uploads to vendor clouds. *Notes / How to Compute:* identify brief upstream bursts from the device/phone to known smart-scale IP/ASNs; count per day.

D7. Weigh-in time variability: Irregular weigh-in times can accompany rapid change [10]. *Notes / How to Compute:* from F6 timestamps, compute circular SD of clock-times over 30 days (requires sufficient F6 events).

Table B.4: Criterion 4 Feature Mapping – DSM-5 MDD ("Insomnia or hypersomnia nearly every day").

Behavior Observation Metric (BOM)	Network-Traffic Feature (ID)	Feasible w/ header & metadata	Detail Ref.
(BOM1) Shifted Sleep Timing	F1 Digital sleep-onset time	Yes	D1
	F2 Digital wake-up time	Yes	D2
	F3 Onset-time variability	Yes	D3
(BOM2) Changed Sleep Duration	F4 Main nightly idle-gap length	Yes	D4
	F5 Sleep-duration z-score	Yes	D5
(BOM3) Sleep Fragmentation	F6 Nocturnal micro-session count	Yes	D6
	F7 Mean inter-awakening gap	Yes	D7
(BOM4) Daytime Hypersomnia / Rhythm Flattening	F8 Day-time idle ratio (08–18 h)	Yes	D8
	F9 Night/Day traffic ratio	Yes	D9

Details for Table B.4

D1. Digital sleep-onset time: Last activity before a long quiet period indicates when the person likely fell asleep [46]. *Notes / How to Compute:* in 22:00–04:00 take last packet before $\geq$30min idle; compare to personal median.

D2. Digital wake-up time: First activity after that quiet period indicates waking [46]. *Notes / How to Compute:* first packet after the same idle gap; flag earlier/later than usual.

D3. Onset-time variability: Irregular bedtimes relate to worse mood and sleep regularity [45]. *Notes / How to Compute:* circular SD of onset times over last 14 nights.

D4. Main nightly idle-gap length: Longest quiet stretch approximates sleep duration; passive Wi-Fi shows good agreement with actigraphy [31]. *Notes / How to Compute:* per night, length of longest $\geq$30min idle in night window.

D5. Sleep-duration z-score: Large deviations from one's norm indicate insomnia or hypersomnia [31]. *Notes / How to Compute:* $z = (L_{\text{sleep}} - \mu)/\sigma$ vs 30-day baseline; interpret $|z| > 1.5$.

D6. Nocturnal micro-session count: Many small night bursts suggest awakenings [46]. *Notes / How to Compute:* count $\leq$5min bursts in 01:00–06:00; mark $\geq$3 bursts on $\geq$3 nights/week.

D7. Mean inter-awakening gap: Shorter gaps mean denser fragmentation. *Notes / How to Compute:* mean time between micro-sessions within the main sleep window.

D8. Day-time idle ratio: Lots of quiet time during the day can reflect napping/hypersomnia [58]. *Notes / How to Compute:* fraction of idle minutes in 08:00–18:00; watch workdays vs weekends.

D9. Night/Day traffic ratio: More night than day activity supports insomnia/circadian delay [19]. *Notes / How to Compute:* activity (bytes or active minutes) in 00:00–06:00 $\div$ 06:00–22:00; track trend.

Table B.5: Criterion 5 Feature Mapping – DSM-5 MDD ("Psychomotor agitation or retardation nearly every day").

Behavior Observation Metric (BOM)	Network-Traffic Feature (ID)	Feasible w/ header & metadata	Detail Ref.
(BOM1) Restless Device Switching (agitation)	F1 Wi-Fi/DHCP re-association events / h	Yes	D1
	F2 Mean Wi-Fi dwell-time	Yes	D2
(BOM2) Slowed / Variable Typing Dynamics (retardation)	F3 Median inter-keystroke gap	Partial	D3
	F4 Typing-speed variability (SD)	Partial	D4
(BOM3) "Screen-check" Burstiness	F5 Sub-30 s sessions / day	Yes	D5
	F6 Session-duration Fano factor	Yes	D6

Details for Table B.5

D1. Wi-Fi/DHCP re-associations / h: Frequent re-connects suggest restlessness [59]. *Notes / How to Compute:* per device, count DHCP cycles (and AP re-assoc if available) per hour from router logs.

D2. Mean Wi-Fi dwell-time: Shorter time staying on one AP indicates pacing/handling [59]. *Notes / How to Compute:* measure time between associations; take daily mean/median.

D3. Median inter-keystroke gap (proxy): Slower typical gaps = psychomotor slowing [64]. *Notes / How to Compute:* within chat sessions, use timing of small upstream bursts as typing proxy; take daily median.

D4. Typing-speed variability (proxy): More variable tempo often accompanies depression [32]. *Notes / How to Compute:* SD of the proxy gaps from D3; require minimum burst count.

D5. Sub-30s sessions / day: Many very short sessions reflect "checking" agitation [46]. *Notes / How to Compute:* sessionize with 5min idle gap; count sessions with duration $<$ 30s.

D6. Session-duration Fano factor: Burstier day = higher variance vs mean [45]. *Notes / How to Compute:* compute $F = \mathrm{Var}(L)/\mathbb{E}[L]$ over session lengths L per day; flag high F vs baseline.

Table B.6: Criterion 6 Feature Mapping – DSM-5 MDD ("Fatigue or loss of energy nearly every day").

Behavior Observation Metric (BOM)	Network-Traffic Feature (ID)	Feasible w/ header & metadata	Detail Ref.
(BOM1) Extended Day-time Inactivity	F1 Longest idle gap 12–18h	Yes	D1
	F2 Day-active session count	Yes	D2
	F3 Day-time idle ratio	Yes	D3
(BOM2) Decline in Effortful Interaction	F4 Upstream bytes / active hours	Yes	D4
	F5 Interactive upstream bursts / day	Partial	D5
(BOM3) Slowed Browsing Tempo	F6 Median inter-request interval	Partial	D6
	F7 Inter-request Fano factor	Partial	D7
(BOM4) Delayed Morning Activation	F8 First-activity time	Yes	D8
	F9 Activation delay vs baseline	Yes	D9

Details for Table B.6

D1. Longest midday idle gap: Long quiet stretches at midday reflect inactivity [57]. *Notes / How to Compute:* in 12:00–18:00, take the longest $\geq$30–40min idle gap; flag if $>$ personal P75.

D2. Day-active session count: Fewer daytime sessions suggest low energy [45]. *Notes / How to Compute:* sessionize traffic (5min idle gap); count sessions starting 09:00–18:00.

D3. Day-time idle ratio: Share of the workday with no activity [57]. *Notes / How to Compute:* fraction of idle minutes in 09:00–18:00; values $>$ 0.4 on workdays are notable.

D4. Upstream bytes per active hour: Less data sent during active hours can indicate reduced effort [60]. *Notes / How to Compute:* sum upstream bytes in hours with any traffic ÷ number of active hours.

D5. Interactive upstream bursts / day (proxy): Fewer small "send" bursts imply less expressive activity [45]. *Notes / How to Compute:* count short, small upstream bursts in chat/productivity domains; use personal 25th percentile as a low-energy marker.

D6. Median inter-request interval (proxy): Slower browsing tempo aligns with fatigue [57]. *Notes / How to Compute:* within browsing sessions, take median time between DNS/SNI touches (or connection starts if DNS hidden).

D7. Inter-request Fano factor (proxy): Erratic bursts with long pauses reflect fluctuating energy [60]. *Notes / How to Compute:* $F = \mathrm{Var}(\Delta t)/\mathbb{E}[\Delta t]$ over inter-request gaps; flag sustained $F > 1.5$.

D8. First-activity time: Later first activity indicates delayed start to the day [57]. *Notes / How to Compute:* earliest packet after 04:00; compare to personal median (+60min = delayed).

D9. Activation delay vs baseline: Persistent delays signal a low-energy phase [45]. *Notes / How to Compute:* $\Delta t = t_{\text{first}} - \text{median}(28\text{d})$; flag if $\Delta t > 60$ min for ≥ 3 consecutive days.

Table B.7: Criterion 7 Feature Mapping – DSM-5 MDD
("Feelings of worthlessness or excessive/inappropriate guilt nearly every day").

Behavior Observation Metric (BOM)	Network-Traffic Feature (ID)	Feasible w/ header & metadata	Detail Ref.
(BOM1) Engagement with self-evaluative or mental-health resources	F1 Mental-health site visits / day	Yes	D1
	F2 "Negative-self" search-query ratio	No	D2
	F3 Time on self-assessment pages	Partial	D3
	F4 Help-line / therapy contact look-ups	Yes	D4
(BOM2) Digital self-withdrawal / data purge behavior	F5 Account-delete / unsubscribe requests	No	D5
	F6 Settings / privacy-page dwell time	Partial	D6
	F7 Decreasing Outgoing-post share vs Reading	Partial	D7
	F8 Decreasing Cloud-upload Volume	Yes	D8

Details for Table B.7

D1. Mental-health site visits / day: Daily sessions to domains labeled "Mental Health." Increased consultation aligns with periods of worthlessness [14]. *Notes / How to Compute:* map eTLD+1 to IAB; count sessions per day.

D2. "Negative-self" search-query ratio: *Not feasible* without reading query text (encrypted). Prior work links such phrases to depressive language [14]. *Notes:* consider a privacy-safe proxy (e.g., night-time search intensity).

D3. Time on self-assessment pages: *Partial:* dwell can be estimated only at the domain level (paths are hidden). Screeners like PHQ/DASS are common [14]. *Notes / How to Compute:* identify assessment platforms by domain; sessionize to approximate dwell.

D4. Help-line / therapy contact look-ups: Sessions to crisis-lines or therapist directories [14]. *Notes / How to Compute:* curated list of helplines/directories; count sessions per day.

D5. Account-delete / unsubscribe requests: *Not feasible* (needs HTTP methods/-paths). StudentLife-style signals of withdrawal are noted in literature [57]. *Notes:* proxy via sessions to account/help hostnames if distinct.

D6. Settings / privacy-page dwell time: *Partial:* only when settings/privacy live on separable hostnames [45]. *Notes / How to Compute:* if distinct domains exist, estimate dwell via session duration.

D7. Decreasing Outgoing-post share vs Reading: Lower share of "sending" vs "reading" on social platforms matches silent rumination [11]. *Notes / How to Compute:* per social session, compute up/down bytes ratio; track downward trend vs baseline.

D8. Decreasing Cloud-upload Volume: Reduced uploads to Drive/iCloud/Dropbox reflect less self-produced content [60]. *Notes / How to Compute:* sum upstream bytes to cloud hostnames/ASNs; compare moving average to personal baseline.

Table B.8: Criterion 8 Feature Mapping – DSM-5 MDD ("Diminished ability to think or concentrate, or indecisiveness").

Behavior Observation Metric (BOM)	Network-Traffic Feature (ID)	Feasible w/ header & metadata	Detail Ref.
(BOM1) Fragmented focus / frequent task-switches	F1 Median page-dwell < baseline	Partial	D1
	F2 DNS burst-rate / h	Yes	D2
	F3 Notification-triggered micro-sessions	Yes (heuristic)	D3
(BOM2) Indecisive information seeking	F4 Repeated-query ratio	Yes (coarse)	D4
	F5 Query reformulation chain length	No	D5
	F6 Back-navigation percentage	Partial	D6
(BOM3) Cognitive sluggishness / slow response	F7 Time-to-first-click on SERP	Partial	D7
	F8 Median inter-keystroke gap	Partial (proxy)	D8

Details for Table B.8

D1. Median page-dwell < baseline (Partial): Approximates attention by time between a visit to one site and the next top-level site; shorter dwell suggests fragmented focus [27]. *Notes / How to Compute:* per session, measure time from one eTLD+1 to the next; take daily median (intersite, not page-true).

D2. DNS burst-rate / hour (Yes): Rapid clusters of different domains indicate tab-hopping/task switching [57]. *Notes / How to Compute:* count clusters of $\geq$3 distinct eTLD+1 within 60s; rate per hour.

D3. Notification-triggered micro-sessions (Yes, heuristic): Push alert touch followed by a quick peek session reflects interrupt-driven attention [27]. *Notes / How to Compute:* detect FCM/APNs touch then $\leq$30s content burst; count pairs.

D4. Repeated-query ratio (Yes, coarse): Many revisits to search engines within a short window suggest indecision [14]. *Notes / How to Compute:* within 60min, count repeated visits to search eTLD+1 $\div$ total sessions (no query text).

D5. Query reformulation chain length (No): Requires reading query text (encrypted) [14]. *Notes:* use a proxy like "# of search revisits before first external site."

D6. Back-navigation percentage (Partial): Approximates "back-and-forth" by returning to the prior site quickly [57]. *Notes / How to Compute:* share of sequences Site A $\rightarrow$ Site B $\rightarrow$ Site A within 30s (cross-domain only).

D7. Time-to-first-click on SERP (Partial): Delay from landing on results to first external site can reflect slowed processing [60]. *Notes / How to Compute:* time from search-engine arrival to first different eTLD+1.

D8. Median inter-keystroke gap (Partial, proxy): Slower typical typing gaps align with cognitive load/retardation [64]. *Notes / How to Compute:* use timing of small upstream bursts in chats as a typing proxy; daily median (requires sufficient bursts).

Table B.9: Criterion 9 Feature Mapping – DSM-5 MDD ("Recurrent thoughts of death, suicidal ideation, or suicide attempt").

Behavior Observation Metric (BOM)	Network-Traffic Feature (ID)	Feasible w/ header & metadata	Detail Ref.
(BOM1) Crisis-oriented help seeking	F1 Crisis-line domain hits	Yes	D1
	F2 Suicide-method query ratio	No	D2
	F3 Therapy-booking page visits	Partial	D3
(BOM2) Self-harm community engagement	F4 Self-harm forum visits	Yes	D4
	F5 Upstream bytes to self-harm forums	Partial	D5
	F6 Avg. session length on those forums	Partial	D6
(BOM3) Farewell / estate preparation	F7 Will/insurance doc downloads	Partial	D7
	F8 Cloud-backup surge	Yes	D8
	F9 Account-deletion requests	No	D9
(BOM4) Nocturnal rumination spikes	F10 Night-time suicide-query bursts	No	D10
	F11 Night-time negative-search ratio	No	D11

Details for Table B.9

D1. Crisis-line domain hits (Yes): More sessions to helplines often precede SI posts [14]. *Notes / How to Compute:* curated helpline eTLD+1; count sessions/day.

D2. Suicide-method query ratio (No): Needs query text (encrypted) [52]. *Notes:* prefer a proxy like "night-time search intensity."

D3. Therapy-booking page visits (Partial): Domain-level only; path details are hidden [57]. *Notes / How to Compute:* therapist-directory/tele-therapy domains; approximate dwell via session duration.

D4. Self-harm forum visits (Yes): Increased visits correlate with SI trajectories [14]. *Notes / How to Compute:* sessions to self-harm–related communities by domain list.

D5. Upstream bytes to forums (Partial): Higher upload suggests posting (noisy) [4]. *Notes / How to Compute:* within forum sessions compute up/down ratio; track rises vs baseline.

D6. Avg. session length on forums (Partial): Longer dwell reflects deeper engagement [4]. *Notes / How to Compute:* sessionize forums with 10–15 min idle gap; average duration.

D7. Will/insurance doc downloads (Partial): File types are hidden; use portal activity as proxy [14]. *Notes / How to Compute:* sessions to estate/insurance domains with large downstream bursts.

D8. Cloud-backup surge (Yes): Archiving spikes can precede mood crashes [60]. *Notes / How to Compute:* upstream to iCloud/Drive/Dropbox hostnames/ASNs; z-score vs 28d baseline.

D9. Account-deletion requests (No): Requires HTTP methods/paths [57]. *Notes:* optional proxy-sessions to account/help hostnames if distinct.

D10. Night-time suicide-query bursts (No): Needs query text [52]. *Notes:* proxy via clusters of night-time search-engine sessions.

D11. Night-time negative-search ratio (No): Needs query semantics [14]. *Notes:* proxy via night/day ratio of help-seeking domain sessions.

Appendix C
GitHub Repository Overview

The GitHub Repository contains the application that powers the Router-Centric Behavioral Signals prototype used in this thesis.[36]

Live Deployment

Hosted build: https://unisg-nef.streamlit.app/

(credentials on request: stephan.nef@student.unisg.ch); alternatively, deploy locally as described below.

Deployment Options:

Listing C.1: Run locally in Docker

```
cd app
docker compose up --build
# Streamlit on http://localhost:8501
```

Listing C.2: Run locally in a Virtual Environment

```
cd app
python -m venv ..\.venv
..\.venv\Scripts\activate   # on Windows (adjust for macOS/Linux)
pip install -r requirements.txt
streamlit run 00_Home.py
```

Notes: install repo-level requirements if running ancillary scripts (`pip install -r ../requirements.txt`); set `SCAPY_USE_LIBPCAP=no` if libpcap is unavailable (the app reads from files and does not require raw PCAP captures).

Safety Reminder Insights are intended to foster early, empathetic conversations. They are not diagnostic labels and must be interpreted by qualified professionals within an ethical data-governance framework.

GitHub Repository Configuration Touchpoints

See Table C.1.

Table C.1: GitHub Repository Configuration Touchpoints

ID	Path	Role
A1	`app/00_Home.py`	Landing page with short disclaimer and summary of what the dashboard does and does not do
A2	`app/pages/01_Early_Signs_Overview.py`	Parquet loading/normalization, FASL & DSM-Gate computation and rendering.
A3	`app/pages/02_Network_Metrics.py`	Orchestrates per-criterion metric computation and UI views.
A4	`app/pages/03_User_and_Network_Settings.py`	Profiles, IP mapping, upload & partition controls for ETL.
A5	`app/metrics/common.py`	Sessionization (`sessions_from_timestamps`), eTLD+1/domain helpers, formatters.
A6	`app/metrics/base_features.py`	Day-level base record (LateNightShare, LongestInactivityHours, IS/IV, etc.).
A7	`app/metrics/criterion1.py–criterion9.py`	Criterion-specific metric definitions and resolvers (C1–C9).
A8	`app/auto_tune.py`	Optional auto-tuning of triangular memberships and weights.
A9	`app/processed_parquet/<dataset>/`	Five-minute Parquet storage used as the canonical on-disk interface between ETL and analytics. Files follow the deterministic pattern `<dataset>__YYYYMMDD_HHMM.parquet` (contiguous 5-minute slices).

Declaration of Authorship

‘I hereby declare,

- that I have written this thesis independently;
- that I have written the articles and contributions using only the aids listed in the index;
- that all parts of the thesis produced with the help of aids (incl. AI-Tools) have been precisely declared;
- that I have mentioned all sources used and cited them correctly according to established academic citation rules; this also includes the comprehensible disclosure of all personal publications;
- that I have acquired all immaterial rights to any materials I may have used, such as images or graphics, or that these materials were originally created by myself;
- that I am aware of the legal provisions regarding the publication and dissemination of parts or the entire thesis and that I comply with them accordingly;
- that I am aware that the University will prosecute a violation of this Declaration of Authorship and that disciplinary as well as criminal consequences may result, which may lead to expulsion from the University or to the withdrawal of my title.’

By submitting this thesis, I confirm through my conclusive action that I am submitting the statutory declaration, that I have read and understood it, and that it is true.

St. Gallen, October 31, 2025

Signature: Nef, Stephan

Declaration of Auxiliary Aids

The creation of this dissertation involved third-party software, including AI-based tools. I declare all uses of third-party software in the dissertation text that contributed non-trivially to the dissertation's content and are non-standard in such research. Further, by the University of St. Gallen's decree II.B.1.20 section 5.3, I explicitly declare the following uses:

- Private proofreading, DeepL, Writefull and OpenAI ChatGPT (Plus, GPT-4o, GPT-5), in their current versions available between February 2025 and October 2025 applied over the whole dissertation, have been used for spelling correction, grammar improvement, style-focused rewriting/paraphrasing and enhancing the conciseness and clarity of the dissertation text.
- OpenAI ChatGPT (Plus, GPT-4o, GPT-5) in their current versions available between February 2025 and October 2025, have been used for brainstorming, code assistance and refactoring and LaTex illustrations, algorithms and formulas.