Imprint

A Literary Portrait from Your Own Words

Technical White Paper

Architecture, Methodology, and Research Foundations

Version	2.0
Date	April 2026
Author	Jack Stone
Contact	jack@myimprint.ai
Website	myimprint.ai

Contents

1. Executive Summary

Imprint is a consumer product that reads a person's messaging history and returns a literary self-portrait structured as ten cards. Not a personality test. Not a horoscope. A hero's journey, grounded in real psychological research, computed from the reader's actual writing, with every claim backed by a verbatim quote they wrote themselves.

The reading combines open-source natural language processing tools (sentiment analysis, emotion detection, topic modeling) with large language model interpretation to produce a portrait that is grounded in quantitative evidence and rendered in literary prose. The goal is not to categorize the reader. The goal is to make them feel **seen** and **sent forward**.

The product addresses a fundamental gap in the self-knowledge market: personality assessments today rely on self-reported questionnaires, which are subject to social desirability bias and limited by conscious self-perception. Imprint inverts this model. Rather than asking users who they think they are, it reads what they have actually written to the people closest to them and reflects it back.

Two reading types share one engine. The Long Imprint (lifetime reading) is produced on first run from the full messaging corpus and serves as the baseline. The Annual Imprint (year in words) is produced each year, compared against the stored baseline, and designed as the retention mechanism. The baseline comparison is the product's structural moat: no competitor can replicate it without having been there the whole time.

Imprint's privacy architecture ensures that intimate claims about a person's inner life never require exposing their private communications. Contact identities are replaced with opaque hashes before any analysis begins. The hash map never leaves the device. The only external API call receives anonymized samples and structural statistics.

Every reading ends with a charge: a single imperative line that tells the reader what to do next. If the reader closes the tab thinking "neat," the reading has failed. If they want to screenshot three cards and take a specific action in the next 24 hours, it worked.

2. Product Architecture

2.1 System Overview

Imprint operates as a five-layer pipeline that transforms raw messaging data into a rendered literary portrait. Each layer has a single responsibility, a defined input/output contract, and an explicit privacy boundary. The layers are: Ingest, Anonymize, Structural Analysis, Interpretive Analysis, and Render.

2.2 Two-Reading System

The Long Imprint consumes the entire available corpus (typically one to ten years of messaging history) and produces the baseline portrait. It answers the question: across all this time, who have you been? The Annual Imprint consumes one calendar year of messaging and compares it against the baseline. It answers: how has this year changed you? The comparison creates the product's most powerful card (Card 08: Who You Are Becoming), where the reader sees their own vocabulary shift over time.

2.3 The Baseline Mechanism

After the Long Imprint renders, the system writes a baseline file to the user's device. This is not a copy of the corpus. It is a compressed statistical fingerprint: typical word frequencies, message-length distributions by hour, attachment-signal weightings, top recurring themes with rolling counts, top contacts and their historical shares, trait scores, peak-hour distributions, and first-mention timestamps.

Each annual run analyzes only the new year's data, then compares against the baseline. The comparisons are the soul of the Annual Imprint. After each annual run, the baseline updates, weighted toward recent years but not erasing deep history. Annual runs may not re-analyze earlier years' raw text. The past is represented only by the baseline statistics.

2.4 Supported Platforms

The current implementation supports iMessage extraction via the macOS chat.db SQLite database. The extraction module handles Apple's nanosecond epoch timestamps, tapback reaction filtering, group chat disambiguation, and contact identity hashing. The architecture is designed to support additional platforms (WhatsApp, Signal, SMS, email) through a common CSV interchange format that normalizes message text, direction (sent/received), timestamp, and contact identifier.

2.5 Minimum Viable Corpus

Twelve months of outgoing messages. Anything shorter is too thin to support the framework. If the corpus is under twelve months, the system tells the user the data is too thin for a full reading and offers a partial reading with a note about which dimensions could not be supported.

3. The Voice

Imprint sounds like a friend who has been watching you closely all year, who is better-read than you are, who loves you, and who will not bullshit you. Not a psychologist. Not a horoscope. Not a corporate wellness app. A friend at a kitchen table at 1am who just put down a glass of wine and said *okay. here's what I've noticed.*

3.1 The Two Things Every Imprint Must Do

Make the reader feel seen. Specific, warm, undeniable. The reader should think "how did it know that." Achieved through real quotes from their own writing and patterns they have never consciously named. Never through clinical labels or percentile scores.

Send the reader forward. Every card ends with a benediction that hands the reader something, not describes them. Every reading ends with a charge. The reader should close it wanting to call someone, write something, start something, say yes to something, or say no to something. If the reader feels studied but not moved, the reading has failed.

3.2 Voice Rules

Second person, always. "You" not "the user." Never the reader's name inside card body copy; only the masthead greeting uses the name.

Short sentences that land like punches. Long sentences when the moment needs to breathe. Never medium sentences in a row.

Specifics beat adjectives. Never "you are caring." Always "you checked on Priya the week her mother got sick and then four more times across the next month."

Numbers as punctuation, not evidence. A number in a card is there to make a line hit harder, not to prove a claim. "2,418 times" lands. "Percentile 84 on openness" dies.

Quotes as moments, not citations. A quote is something the reader said, not data we extracted. No timestamps on quotes unless the timestamp is part of the emotional point.

No academic attribution in the card frame. No "after Pennebaker." No "after Bowlby." The intellectual lineage belongs in this paper, not on the card.

No confidence indicators in the card. We are either confident enough to make the claim or we cut the claim. Showing a three-out-of-five confidence dot undermines emotional force.

No em-dashes. Ever. Use periods, commas, colons, or line breaks. Zero tolerance.

3.3 Benedictions

Every card ends with a benediction. The benediction is the moment the card stops describing the reader and starts speaking to them. It is direct address. It hands them something forward. Not a summary. A gift. The benediction lives after the evidence and the body of the card, set apart visually.

3.4 The Energy

The reader should leave fired up, but Imprint is not a hype video. The energy is the energy of a really good commencement address, a really good letter from a mentor, the last page of a novel that rearranged you. It earns the emotion through specificity and then lets the final line launch the reader without shouting at them.

> *Good: "Stop waiting for someone to hand you the room. You already own it. The last six months of your writing are the receipt."*
>
> *Bad: "CRUSH YOUR GOALS. YOU ARE UNSTOPPABLE. LIVE YOUR BEST LIFE."*

4. The Five-Layer Privacy Pipeline

Layer 1: Ingest

Reads the raw messaging database and produces a normalized CSV. For iMessage, this involves querying macOS chat.db via SQLite, converting Apple epoch timestamps (nanoseconds since January 1, 2001) to ISO 8601, filtering tapback reactions (message types 2000-2005), removing empty messages and attachment-only entries, and extracting chat identifiers for group vs. individual thread disambiguation. Output: a normalized CSV with columns for timestamp, text, contact identifier, direction, and thread type.

Layer 2: Anonymize

Replaces all contact identifiers with deterministic SHA-256 hashes. The hash map (contact_map.json) is stored separately and never leaves the device. All downstream processing operates exclusively on hashed identifiers. This means the structural analysis, interpretive analysis, and rendered output contain zero personally identifiable information about the reader's contacts.

Layer 3: Structural Analysis

Applies three open-source NLP tools to the anonymized corpus: VADER (Valence Aware Dictionary and sEntiment Reasoner) for sentiment polarity scoring, NRC Emotion Lexicon for eight-dimensional emotion classification, and Empath for 200-category lexical topic modeling. Additionally computes: word frequency distributions with first-mention timestamps, hourly and daily volume patterns, message length statistics, contact-level engagement metrics, monthly emotion and topic profiles, longitudinal vocabulary drift analysis, laughter escalation counts, linguistic style markers (question frequency, exclamation frequency, first-person pronoun rate, hedging language), and "weather" identification: calm weeks vs hard weeks (VADER compound below personal baseline for 7+ consecutive days). Output: structural_analysis.json.

Layer 4: Interpretive Analysis

Passes the structural analysis and curated corpus samples through a multi-pass large language model pipeline (Claude API). Ten carefully constructed prompts, one per dimension, each grounded in the research tradition for its card, produce scored claims with verbatim evidence quotes and internal confidence scores. Evidence verification is non-negotiable: every quote returned gets exact-match verified against the original hashed corpus before anything renders. Any claim whose quote cannot be verified character-for-character is dropped. This is the hallucination firewall. Output: cards.json containing all card content, type name, and narrative metadata.

Layer 5: Render

Populates an HTML template with the card content from cards.json. Injects accent color CSS variables based on the reader's dominant emotional texture. Generates SVG visualizations (constellation, orbit, wave, arc) from card data. Adds scroll-triggered count-up animations and fade-in transitions. The device rehydrates hashed tokens into real names by looking them up in the local hash map. The user sees real names on their screen because the names live on their machine. The server only ever sent back hashed tokens. Output: a self-contained HTML file that is the reader's Imprint.

5. Computational Linguistics Methodology

5.1 VADER Sentiment Analysis

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool specifically attuned to social media and informal text. It produces four scores per text unit: positive, negative, neutral, and a compound score normalized between -1 (most negative) and +1 (most positive). VADER handles common social media conventions including emoticons, slang, acronyms (e.g., LOL), and punctuation-based emphasis (e.g., exclamation marks).

In the Imprint pipeline, VADER is applied at both the message level and the aggregated level. Message-level scores are used to identify sentiment valleys (hard weeks) and peaks (high-vitality periods). Monthly compound score trajectories provide the primary data for Card 06 (Your Weather) and Card 09 (The Turn), where sentiment reversals indicate potential turning points.

> *Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.*

5.2 NRC Emotion Lexicon

The NRC Emotion Lexicon (EmoLex) associates words with eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) and two sentiments (positive, negative). The lexicon contains approximately 14,000 English words, each manually annotated via crowdsourcing. Imprint uses NRC to construct monthly emotion profiles: the relative proportion of each emotion in the reader's writing for a given month.

These emotion profiles feed Card 02 (The Shape of You), where emotion distributions are mapped to personality trait approximations, and Card 05 (What Makes You Come Alive), where high-joy and high-anticipation periods are correlated with specific topics and contacts to identify the reader's primary vitality source.

> *Mohammad, S.M. & Turney, P.D. (2013). Crowdsourcing a Word-Emotion Association Lexicon. Computational Intelligence, 29(3), 436-465.*

5.3 Empath Topic Modeling

Empath is a tool for analyzing text across 200 pre-built lexical categories (e.g., family, achievement, violence, optimism, cooking, travel). Unlike topic models such as LDA, Empath uses a neural embedding approach: categories are defined by seed terms, and the lexicon is expanded by finding semantically similar words in a large corpus. This makes it robust to the informal, fragmented text typical of messaging.

Empath provides the primary data for Card 07 (What You Keep Returning To), where topic recurrence ratios identify the reader's three most persistent preoccupations. It also feeds the accent color selection algorithm, which assigns one of four color palettes (ocean, amber, plum, sage) based on the dominant

topical texture of the corpus.

Fast, E., Chen, B., & Bernstein, M.S. (2016). Empath: Understanding Topic Signals in Large-Scale Text. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 4647-4657.

5.4 Supplementary Analyses

Beyond the three primary NLP tools, the structural analysis layer computes several additional metrics drawn from computational linguistics research. Word frequency analysis with first-mention timestamps enables signature pattern detection (Card 01: Your Signal). Pennebaker's function-word framework informs the analysis of pronoun usage, verb tense distribution, and article frequency as indicators of psychological state. Message-length distributions and hourly volume patterns provide behavioral fingerprints. Longitudinal vocabulary drift analysis (earliest vs. latest quartile word frequency comparison) identifies emerging and fading terms for Card 08 (Who You Are Becoming).

6. The Imprint Framework: Ten Dimensions

The Imprint Framework organizes the reading into ten dimensions structured as a three-act hero's journey. Each dimension is grounded in a specific research tradition, mapped onto the corpus through computational methods, and rendered as a single card with one central claim, supporting evidence, and a closing benediction. The three-act structure creates emotional pacing: Act I reflects the reader's observable patterns, Act II reveals what lies underneath those patterns, and Act III projects where the reader is headed.

Act I: The Mirror

Who you have been.

01 : Your Signal

The reader's linguistic fingerprint. Leads with the single most striking count in the corpus (e.g., a distinctive word used 9,847 times). Traces the signature pattern across time. Grounded in Pennebaker's function-word analysis and message-length distributions.

Research: Pennebaker (2011)

02 : The Shape of You

Big Five personality traits rendered as a constellation, not a dashboard. Traits are inferred from writing patterns using validated linguistic correlates (vocabulary diversity for openness, inclusive pronouns for agreeableness). The two brightest traits are named in human language.

Research: Big Five / HEXACO; Yarkoni (2010)

03 : Your Tell

The single most statistically distinctive behavior in the corpus. One hero number, one claim, maximum whitespace. This is the screenshot card, designed for sharing. The tell must be specific enough that two-thirds of people would say "that's not me."

Research: Statistical outlier detection

Act II: The Truth

What is underneath.

04 : How You Love

The reader's attachment pattern rendered in a single sentence, backed by contact-level messaging data. An orbit visualization shows the reader at center with close contacts as satellites. Attachment style is inferred from response latency curves, emotional-tone shifts by recipient, top-contact stability over time, and behavior changes toward close contacts during stress.

Research: Bowlby (1969); Ainsworth et al. (1978); Hazan & Shaver (1987)

05 : What Makes You Come Alive

The reader's primary vitality source identified through self-determination theory. High-vitality periods are detected via co-occurrence of positive emotion words, increased message length, expanded vocabulary, and exclamation frequency. The dominant psychological need (autonomy, competence, or relatedness) is inferred from which Empath category clusters peak during high-vitality windows.

Research: Deci & Ryan (1985); Self-Determination Theory

06 : Your Weather

How the reader behaves when things get hard. Hard weeks are identified as consecutive days where VADER compound scores fall below the reader's personal baseline. Behavioral shifts during these periods (message length changes, vocabulary contraction, humor frequency, self-reference patterns) are compared against baseline to identify the stress response mechanism. The claim is always framed as a form of strength.

Research: VADER sentiment valleys; linguistic style markers

Act III: The Horizon

Where you are going, whether you have admitted it or not.

07 : What You Keep Returning To

Three recurring themes identified through Empath topic-cluster analysis with rolling frequency counts. Themes are named as questions or preoccupations, not categories. Selection criteria: highest recurrence-to-total-volume ratios, filtered for topical distinctiveness (excluding universal topics). Each theme includes ordinal, frequency count, first-mention date, and a verbatim quote.

Research: Empath topic recurrence analysis

08 : Who You Are Becoming

Longitudinal vocabulary drift rendered as a then/now reveal. In Long mode, compares earliest quartile vs. latest quartile of the corpus. In Annual mode, compares January vs. December. The single most telling word shift is highlighted. Grounded in McAdams's narrative identity framework: identity is a story that changes as the person develops.

Research: McAdams (1993); narrative identity

09 : The Turn

The single biggest turning point in the corpus, identified via sentiment reversal detection (VADER) combined with topic shift analysis. The card presents a long verbatim blockquote from the moment, framed by context and interpretation. Grounded in McAdams's redemption sequences: moments where suffering is reframed as growth.

Research: McAdams (2006); redemption sequences

The Reveal

10 : Your Imprint

The culminating card. Dark background, inverted palette, full-bleed. A literary type name (freshly minted, two to six words, image-rich), a synthesis paragraph that weaves threads from all nine prior

cards, and a charge (a single imperative line designed to make the reader act within 24 hours). Grounded in McAdams's concept of the personal myth: the narrative identity a person constructs to make sense of their life.

Research: McAdams (1993); personal myth

7. Psychological Research Foundations

Each dimension of the Imprint Framework is grounded in established psychological research. This section details the six primary research traditions that inform the product, how each is operationalized through computational methods, and the limitations of each mapping.

7.1 Pennebaker's Function-Word Analysis

James Pennebaker's research program, spanning three decades, demonstrates that the small, seemingly insignificant words people use (pronouns, articles, prepositions, auxiliary verbs) reveal more about their psychological state than the content words they choose. People who use more first-person singular pronouns ('I', 'me', 'my') tend to be more self-focused and are at higher risk for depression. People who use more first-person plural pronouns ('we', 'us', 'our') tend to have stronger social integration. Verb tense usage reflects temporal orientation: past-tense dominance correlates with depression, while future-tense usage correlates with goal-directed behavior.

In Imprint, function-word analysis informs Card 01 (Your Signal) and contributes to Card 02 (The Shape of You). The pipeline counts first-person pronoun frequency, question frequency, exclamation frequency, and message-length distributions. These are not used to diagnose but to characterize: to name the shape of the reader's voice in specific, quantitative terms.

Limitation: Pennebaker's research is based primarily on extended writing samples (essays, journals). Messaging is shorter, more fragmented, and more context-dependent. Imprint accounts for this by aggregating across thousands of messages rather than analyzing individual texts.

7.2 Big Five Personality Traits

The Five-Factor Model (openness, conscientiousness, extraversion, agreeableness, neuroticism) is the most empirically validated personality taxonomy in psychology. Yarkoni (2010) demonstrated that Big Five traits can be predicted from writing samples with moderate accuracy using linguistic features: vocabulary diversity and abstract language correlate with openness; inclusive pronouns and positive emotion words correlate with agreeableness; social references and exclamation marks correlate with extraversion.

Imprint approximates Big Five scores by mapping Empath topic category activations and VADER sentiment distributions onto validated linguistic correlates. The resulting scores are presented as a constellation visualization, not as percentile rankings, to avoid the clinical framing that makes personality assessments feel reductive. The card names traits in human language ('warmth' rather than 'agreeableness') and focuses on the two most dominant traits.

Limitation: Personality inference from text is approximate. Imprint presents trait scores as relative patterns within the reader's own corpus, not as validated psychometric measurements. The product explicitly avoids clinical language and percentile comparisons.

7.3 Attachment Theory

Attachment theory (Bowlby, 1969; Ainsworth et al., 1978) proposes that early caregiving experiences create internal working models of relationships that persist into adulthood. Hazan and Shaver (1987) extended this framework to adult romantic relationships, identifying three primary styles: secure, anxious-preoccupied, and dismissive-avoidant. Each style produces distinctive communication patterns: secure individuals show consistent emotional availability; anxious individuals show heightened contact-seeking during stress; avoidant individuals show withdrawal during stress.

Imprint maps attachment patterns through contact-level messaging data: response latency curves, emotional-tone shifts by recipient, the stability of the top-contact list over time, and behavioral changes toward close contacts during identified stress periods. The resulting pattern is described in human language ('You are the one who stays') rather than clinical categories. The card explicitly names both the strength and the cost of the pattern.

Limitation: Attachment style cannot be reliably diagnosed from messaging data alone. Imprint describes observable messaging patterns and frames them through an attachment lens without making clinical claims. The sensitivity rule requires that the card never pathologize and always name the strength in the pattern before the cost.

7.4 Self-Determination Theory

Self-Determination Theory (Deci & Ryan, 1985) identifies three basic psychological needs: autonomy (the need to feel volitional), competence (the need to feel effective), and relatedness (the need to feel connected). When these needs are satisfied, people experience vitality, intrinsic motivation, and well-being. When they are thwarted, people experience depletion and disengagement.

Imprint identifies high-vitality periods in the corpus through co-occurrence of positive NRC emotion scores, increased message length, expanded vocabulary diversity, and elevated exclamation frequency. The dominant SDT need is inferred from which Empath category clusters peak during these windows: achievement and work categories suggest competence; social and family categories suggest relatedness; freedom and travel categories suggest autonomy.

7.5 McAdams's Narrative Identity

Dan McAdams's research on narrative identity proposes that people construct internalized, evolving life stories that integrate the reconstructed past and imagined future to provide life with unity and purpose. Two concepts from this framework are central to Imprint: the personal myth (the overarching story a person tells about who they are) and redemption sequences (moments where suffering is reframed as growth or learning).

The personal myth informs Card 10 (Your Imprint), where the type name crystallizes the narrative identity observed across the corpus. Redemption sequences inform Card 09 (The Turn), where sentiment reversal detection identifies moments of reframing. Narrative identity also informs Card 08 (Who You Are Becoming), where vocabulary drift over time reveals how the reader's self-story is evolving.

7.6 Pennebaker's Linguistic Inquiry

Beyond function words, Pennebaker's broader research program (culminating in the LIWC tool) demonstrates that linguistic style markers track psychological processes in real time. People who are processing trauma use more cognitive mechanism words (because, reason, think). People experiencing social integration use more inclusive language (we, together, with). People who are being deceptive use fewer self-references and more negative emotion words.

While Imprint does not use LIWC directly (it relies on the open-source VADER, NRC, and Empath stack), the conceptual framework of linguistic inquiry informs the pipeline's design: the conviction that writing style, analyzed at scale, reveals psychological patterns that the writer may not consciously recognize.

8. The Interpretive Layer: Large Language Models

8.1 Why an LLM Layer Is Necessary

The structural analysis layer produces quantitative findings: sentiment scores, emotion profiles, word frequencies, topic weights. These findings are necessary but not sufficient for a reading that resonates emotionally. The gap between 'your positive sentiment peaked in March' and 'you fell in love in March and every sentence got three words longer' is the gap that the interpretive layer bridges.

Specifically, the LLM layer provides capabilities that statistical NLP tools cannot: sarcasm and irony detection (VADER reads 'great, just great' as positive; the LLM understands it is not); relationship dynamic inference (the LLM can read a thread and identify someone pulling away or managing a conflict); meaning reversal detection (the moment someone starts using 'we' instead of 'I', or stops qualifying every statement); and literary prose generation (the voice that makes the reading feel like it was written by someone who knows the reader).

8.2 Multi-Pass Architecture

The interpretive layer uses a five-pass architecture designed to maximize output quality while maintaining narrative coherence across all ten cards.

Pass 1: The Scan

A single Claude Sonnet call that receives the full structural analysis and a representative sample of corpus messages. Its task is to identify the five most interesting patterns and draft a narrative brief (400-600 words) that names the story arc of the reading. This brief becomes the connective tissue shared with all subsequent passes, ensuring that the ten cards tell one coherent story rather than ten isolated observations.

Pass 2: Card Generation

Nine parallel Claude Sonnet calls, one per card (Cards 1-9). Each call receives: the dimension-specific prompt (grounded in the research tradition for that card), the narrative brief from Pass 1, the subset of structural analysis data relevant to that card, and curated corpus samples. Parallelization reduces generation time from minutes to seconds.

Pass 3: The Imprint

A single Claude Opus call that generates Card 10. This is the only card that uses the more capable model, because it must synthesize the entire reading into a type name, narrative paragraph, and charge. It receives the complete output from all nine prior cards plus the structural analysis summary. The type name must reference specific details from earlier cards to feel earned.

Pass 4: Coherence Check

A single Claude Sonnet call that reviews all ten completed cards together. It checks for: narrative arc consistency (do the cards tell one story?), redundancy (do any two cards make the same observation?), specificity (could any claim apply to anyone?), cross-references (does Card 10 reference earlier cards?), and voice consistency (any slips into clinical tone?). Returns either 'ship' or specific revision instructions per card.

Pass 5: Targeted Revisions

If the coherence check flags specific cards, those cards are regenerated with the editor's feedback injected into the prompt. Only flagged cards are revised; the rest are preserved. This pass runs zero to three calls depending on the coherence check results.

8.3 Cost Structure

At current API pricing, the interpretive layer costs approximately $0.60-0.80 per reading. Pass 1 (Sonnet scan): ~$0.03. Pass 2 (9x Sonnet cards): ~$0.35. Pass 3 (Opus Card 10): ~$0.15. Pass 4 (Sonnet coherence): ~$0.04. Pass 5 (revisions, if needed): ~$0.08. This cost structure supports a consumer price point of $4.99-9.99 per reading with healthy margins.

8.4 Prompt Engineering

Each of the ten card prompts is a standalone document (800-1500 words) that specifies: what the card does, what research tradition grounds it, what data it consumes, what the output schema looks like, what good output looks like (with examples), what bad output looks like (with examples), and which horoscope traps to avoid. The horoscope trap framework is critical: it defines the boundary between a claim that is specific to this reader and a claim that could apply to anyone. Every prompt includes explicit examples of both.

9. Rendering and Design

9.1 Palette

Core palette: cream (#F4EFE3), terracotta (#B8543F), ink (#1A1614). Typography: Fraunces (display), Newsreader (body), JetBrains Mono (stats). Consistent with myimprint.ai.

9.2 Accent Color System

Every reading picks a secondary accent color from the dominant texture of the person's life. This is not decoration. It is the product's way of saying *this one is yours.*

Water, sky, travel: ocean navy (#2C5F7E). Rooms, cooking, warmth: warm amber (#C7954A). Nights, cities, edge: deep plum (#5B3A6B). Earth, outdoors, steadiness: sage (#6B7F5E).

The accent is assigned at first run and persists across all future readings for this user. It appears on cards where it fits the emotional register (How You Love, What Makes You Come Alive, Becoming, the final Imprint card). Cards about the body of the year (Signal, Tell, Turn) stay in core terracotta. The two palettes live together; the accent never replaces terracotta, it joins it.

9.3 Numbers as Texture

Spotify Wrapped density. Every card has micro-stats sprinkled in small mono font: counts, percentages, time-of-day, word drops, word spikes, contact shares. Numbers are grace notes, not the main event. They earn their place by being specific enough to be unforgettable (2,847, not "a lot"; 67% to six people, not "most to close friends"). The hero number of each card gets big display treatment. Supporting numbers live in stat strips.

9.4 Typography

The primary visual element. Most cards are 70% typography, 30% everything else. Big display type for emotional beats. Small mono for grace notes. Never more than two type families active on one card.

9.5 Graphics

Feelings, not measurements. Bloom, constellation, rising line, door, threshold, spiral, horizon, wave, orbit. Never a bar chart. Never a percentile. Never a quadrant plot with axis labels.

9.6 Motion

Ambient, not theatrical. Count-up reveals on numbers when scrolled into view. Slow drift on the constellation. Stroke draw-in on the arc. Gentle twinkle on star points. Slow pulse on the final charge. All motion respects prefers-reduced-motion. The test: if a reader turns off motion, the reading is still

complete and still moves them.

9.7 Whitespace

Cards should have room to breathe, especially around the emotional peak. The peak is usually the hero line or the final benediction. Give it a clear stage.

9.8 Structural Rules

No confidence meters on cards. No citation bylines on cards. No act numbers in card headers (act titles appear only in section dividers). Card 10 is a dark card (inverted palette, full-bleed).

10. Privacy and Security Architecture

10.1 Design Principles

Imprint's privacy architecture is built on three principles: (1) minimize data movement (process on-device whenever possible), (2) anonymize before analysis (no personally identifiable contact information enters the analytical pipeline), and (3) discard after rendering (the raw corpus is not retained after the reading is generated).

10.2 Contact Identity Protection

All contact identifiers (phone numbers, email addresses, chat handles) are replaced with deterministic SHA-256 hashes at the anonymization layer. Given a hash, it is computationally infeasible to recover the original identifier. The mapping between hashes and identifiers (contact_map.json) is stored locally and never transmitted. All downstream processing, including the Claude API calls, operates exclusively on hashed identifiers. This means the interpretive layer can reference 'contact_0a3f' without ever knowing who that person is.

10.3 Corpus Sampling for API Calls

The interpretive layer does not send the entire corpus to the Claude API. Instead, a curated sampling layer selects approximately 500 representative messages using six sampling strategies: longest messages (richest interpretive material), extreme sentiment messages (highest and lowest VADER scores), top-contact messages (most frequent relationships), temporal extremes (earliest and latest quartiles), questions (reveal preoccupations), and high-volume day messages (indicate significant events). All samples are anonymized before transmission.

10.4 Data Lifecycle

After the reading is rendered, the pipeline produces five output files: the anonymized CSV, the contact hash map (private), the structural analysis JSON, the interpreted cards JSON, and the rendered HTML. The raw messaging database (chat.db) is never copied or modified. The user controls all output files and can delete them at any time. No data is transmitted to Imprint servers; the only external API call is to the Claude API for interpretation, which receives only anonymized samples and structural statistics.

11. Validation and Quality Assurance

11.1 The Horoscope Test

Every claim in an Imprint reading must pass the horoscope test: if a claim could apply to more than one-third of the general population, it fails. 'You care deeply about your loved ones' fails (everyone believes this). 'You ended 9,847 sentences with ha. across four years' passes (this is specific to one person). The horoscope test is enforced at two levels: in the per-card prompts (which include explicit examples of horoscope failures) and in the coherence check pass (which flags generic claims for revision).

11.2 Quote Verification

Every verbatim quote attributed to the reader must be verified by exact-match lookup against the original corpus. Hallucinated quotes are a product killer: if the reader encounters a quote they do not recognize, the entire reading loses credibility. The render layer includes a verification step that flags any quote not found in the corpus samples.

11.3 Sensitivity Guardrails

Two cards carry explicit sensitivity rules. Card 04 (How You Love) must name the strength in the attachment pattern before naming its cost, and must never pathologize. Card 09 (The Turn) must honor genuine loss before interpreting the turning point, and must never manufacture a redemption narrative where the evidence does not support one. If hard weeks correlate with crisis language, Card 06 (Your Weather) produces a gentle acknowledgment rather than an aestheticized analysis of suffering.

11.4 Coherence Verification

The Pass 4 coherence check serves as an automated editorial review. It verifies that the ten cards form a narrative arc (not a list of findings), that no two cards make redundant observations, that Card 10's type name feels earned by the preceding nine cards, and that the voice remains consistent (second person, literary, no academic citations, no em-dashes). Cards that fail the coherence check are revised in Pass 5.

11.5 What the Reader Must Feel

Recognized. Slightly exposed, in a way that is exciting rather than shameful. Moved. Sent forward. Like they just read a letter written specifically to them by someone who gets it.

11.6 What the Reader Must Never Feel

Studied. Graded. Diagnosed. Categorized alongside strangers. Like they are reading about someone else. Like the machine is showing off.

11.7 The Non-Negotiable Rules

1. **Never invent quotes.** Every quoted string must be verified by exact-match lookup. Paraphrasing counts as inventing.

2. **Never assume facts about the user.** The corpus is the only source of truth. Volume drops become "you went quiet," not "you went quiet around your wedding."

3. **Never mix computation and interpretation in the same step.** Run structural analysis first, save to disk, then write cards from the saved output only.

4. **Zero em-dashes or en-dashes in final output.** Grep before delivery.

5. **Handle grief and hardship with care.** Never aestheticize genuine suffering.

6. **Life events require verbatim grounding.** Volume changes alone cannot carry a life-event claim.

7. **The hash map never leaves the device.** This is the privacy guarantee.

8. **No academic attribution in cards.** The research lineage lives in the framework, not on the card.

9. **No confidence indicators in cards.** Confident enough to claim it, or cut it.

10. **Every card has a benediction.** No exceptions.

12. References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation.* Lawrence Erlbaum Associates.

Bowlby, J. (1969). *Attachment and Loss: Vol. 1. Attachment.* Basic Books.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior.* Springer.

Fast, E., Chen, B., & Bernstein, M. S. (2016). Empath: Understanding Topic Signals in Large-Scale Text. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4647-4657.

Hazan, C., & Shaver, P. (1987). Romantic Love Conceptualized as an Attachment Process. *Journal of Personality and Social Psychology*, 52(3), 511-524.

Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.*

McAdams, D. P. (1993). *The Stories We Live By: Personal Myths and the Making of the Self.* William Morrow and Company.

McAdams, D. P. (2006). *The Redemptive Self: Stories Americans Live By.* Oxford University Press.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3), 436-465.

Pennebaker, J. W. (2011). *The Secret Life of Pronouns: What Our Words Say About Us.* Bloomsbury Press.

Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, 55(1), 68-78.

Yarkoni, T. (2010). Personality in 100,000 Words: A Large-Scale Analysis of Personality and Word Use among Bloggers. *Journal of Research in Personality*, 44(3), 363-373.