

# 1. Extension Design Logic

## Comment Monitoring

- Content script detects 2 platforms: Twitter and Instagram.
- Captures user comments when submit button clicked
- Sends to background script for analysis

## AI Classification

- Primary: TensorFlow.js model (runs in offscreen document)
- Fallback: Keyword matching across 6 categories
- Categories: Normal, Profanity, Derogatory, Hate Speech, Microaggression, Trolling

## Sentiment Mapping

- "Normal" category → positive sentiment
- All other 5 categories → negative sentiment
- No actual neutral sentiment (the code only outputs positive or negative)

## Scoring System

- Positive (Normal): +3 health, +3 growth, +3 water drops
- Neutral: +2 health, +2 growth, +2 water drops (but rarely used)
- Negative (toxic): -3 health, -3 growth, +3 poison drops

## Tree Growth Mechanics

- Health: 0-100 (dies at 0)
- Growth Progress: 0-100
- Growth stages by progress:
  - 0-19: Seed 🌱
  - 20-39: Sprout 🌿
  - 40-59: Sapling 🌳
  - 60-79: Young Tree 🌳
  - 80-99: Mature 🌳
  - 100 + health  $\geq$  70: Fruiting 🍎 🍊 🍷 🍒 🍋
- Water/poison drops accumulate (visual counter)
- Tree status: "growing" or "dead"

## Data Persistence

- Local: Chrome storage (immediate)
- Backend: Supabase sync every 10 minutes
- Tracks: extensionUserId, comments, tree stats, weekly stats

# 2. Behavioral Intervention Design and Psychological Framework

## Participant Onboarding and System Initialization

The intervention was delivered via a custom browser extension designed for cross-platform integration. The onboarding process was engineered to establish psychological ownership and ensure data privacy. Upon installation, the system executed a three-stage setup:

1. **Pseudonymization:** The system auto generated a unique Universal Unique Identifier (UUID), which was mapped to the participant's Prolific ID. This ensured that all subsequent behavioral data remained decoupled from the participant's platform-level identity.
2. **Platform Integration:** Participants granted the extension permission to monitor interactions on a two major social media platforms (Twitter/X and Instagram).
3. **Agency and Stewardship:** To foster intrinsic motivation, participants selected one of six tree species (cherry, apple, lemon, peach, coconut palm and orange), each characterized by distinct growth parameters and visual profiles. The intervention commenced with the tree in a "seedling" state (100% health, 0% growth).

## Psychological Rationale for the Tree Metaphor

The choice of an organic, evolving visualization over traditional quantitative metrics (e.g., bar charts or numerical scores) was grounded in four behavioral principles:

- **Concreteness Effect:** Traditional moderation metrics are often abstract. A living organism provides a relatable, high-fidelity representation of social impact that reduces the "Distance from Effect" inherent in digital communication.
- **Cumulative History:** While a "karma score" represents a snapshot in time, a mature tree visually encapsulates weeks of prosocial investment, making the long-term impact of routine behavior salient.
- **Non-Binary Recovery:** Unlike traditional moderation (e.g., "banned" vs. "active"), the health-state model (100% to 0% health) allows for behavioral correction and recovery, supporting the "reflective" rather than "punitive" design goal.
- **Loss Aversion:** Grounded in prospect theory, the visual deterioration of a nurtured artifact triggers a stronger emotional response than the mere reduction of a numerical score, encouraging more cautious expressive choices.

## Feedback Latency and Temporal Dynamics

To support associative learning (linking behavior to consequence), the system operated on a multi-tiered feedback schedule:

- **Immediate Feedback (Latency <2s):** Real-time updates to tree health and growth, accompanied by ambient visual pulses (green for constructive/positive, red for harmful/anti-social), were triggered upon comment submission.
- **Daily and Weekly Synthesis:** Social comparison features (leaderboards) and activity summaries were updated at longer intervals to prevent "gaming" behavior and to encourage sustained, reflective engagement rather than compulsive checking.
- **Synchronous State Persistence:** To maintain a consistent "Distance from Effect" across different browsing sessions, the tree state was synchronized with the backend every 5 minutes, ensuring the persistent visualization remained a stable anchor for the participant.