

## Digital Persona Schema for a Content Creator



Modern AI can turn a creator's past videos into a "**digital persona**" profile by mining dozens of features. In practice, state-of-the-art video analysis pipelines extract low-level visual metrics (colorfulness, shot cuts, face detections, etc.) and semantic cues from transcripts <sup>1</sup> <sup>2</sup>. For example, one study consolidated 109 *video-based variables* – ranging from color palette and scene-cut frequency to presence of faces – all without manual labeling <sup>1</sup>. Likewise, analysts use NLP on transcripts to pull out topics, catchphrases and emotional tone <sup>3</sup>. These methods show it's entirely feasible to *algorithmically encode* a creator's style: the visual tempo of their edits, their average speech rate, favorite buzzwords, tone and so on. (This builds on the idea of a "Digital Author Persona" recently formalized with a JSON schema <sup>4</sup>.)

**Key Techniques:** To extract a persona, we combine:

- **Computer Vision:** frame-by-frame analysis (via CNNs/OpenCV) for shot length, camera angles, colors, graphics and on-screen text <sup>1</sup>.

- **Audio Analysis:** acoustic metrics (speech rate, pitch, energy) and background sound classification (music genre, effects).

- **NLP/LLMs:** transcript processing with topic modeling, sentiment analysis, and pattern mining. For instance, spoken words are converted to text; then a model (GPT-4/Gemini) identifies recurrent topics, catchphrases and narrative structure. Studies confirm that video transcripts carry "valuable properties of context and emotion" useful for recognizing topics and emotional valence <sup>3</sup>.

- **Metadata & Engagement Signals:** extracts structured tags and viewer response metrics. Modern video AI even tags objects/activities and computes an overall emotional tone <sup>2</sup>, which we can fold into the persona.

These multimodal features are merged into a **structured persona profile**. In effect, we treat the creator's channel as a unique "Author Persona" in data form <sup>4</sup>. The goal is a comprehensive JSON that computers

can read, but humans can also understand with comments. This schema can power tools that detect **content drift** ("Has my channel personality shifted?") or even **persona injection** (rewriting content in the creator's own style).

## Persona Schema Outline

Below is a *proposed JSON schema* capturing an account's creative persona across modalities and platforms. Each section includes comments (human-readable notes) explaining the fields:

```
{
  "creator_id": "unique_creator_handle",      // Unique ID or channel name

  "identity": {
    "niche": ["tech_review", "productivity"], // Core topics or niche areas
    "personality_traits": {
      "enthusiasm": 0.85,      // [0-1] scale of energetic tone
      "humor": 0.7,           // [0-1] self-descriptive style metrics
      "seriousness": 0.3
    },
    "self_description": "Tech guru who loves gadgets and lifehacks." // Brand/intro line
  },

  "visual_dna": {
    "camera_angles": ["close_up", "medium_shot"],   // Frequent shot types
    "lighting": "bright",                           // e.g. bright, natural, dramatic
    "color_palette": ["#FF0000", "#000000", "#FFFFFF"], // Dominant hues in thumbnails/videos
    "backgrounds": ["home_office", "on_location"],   // Common background settings
    "graphic_style": {
      "text_overlay": true,
      "transitions": ["cut", "dissolve"],          // Typical edit transitions
      "animation": "minimal",                     // e.g. minimal, moderate, heavy
      "filters": ["vibrance_boost"]                // e.g. color grading filters
    },
    "cut_length_stats": {
      "avg_shot_length_sec": 4.2,                  // Average seconds per shot (rapid or slow edits)
      "cut_rate_per_min": 15                        // Cuts per minute
    },
    "camera_movement": {
      "steadycam": 0.4,                          // [0-1] fraction of time handheld vs fixed
    }
  }
}
```

```

        "zoom_frequency": 0.5                                // times/sec the camera zooms
    },
    "on_screen_elements": {
        "face_presence": true,                            // Does creator appear on
camera?
        "slides_screen_share": false,
        "props": ["laptop","microphone"]                // Recurrent props or set
pieces
    }
},
"audio_dna": {
    "speech": {
        "avg_wpm": 165,                               // Words per minute (speaking rate)
        "monotone": false,                            // Uses varying intonation vs flat
        "pause_rate": 0.1                            // Fraction of time silent/pausing
    },
    "music": {
        "uses_music": true,
        "genre": ["lo-fi","electronic"],   // Typical background music genres
        "volume_norm": 0.3,                  // [0-1] relative volume level
        "intro_jingle": true,                // Custom intro/outro music present
        "sound_effects": ["whoosh","ding"] // Common SFX (e.g., transitions/
notification sounds)
    },
    "audio_tone": {
        "warmth": 0.6,                         // [0-1] e.g. measured by timbre or prosody
        "energy": 0.8,                          // high=energetic delivery
        "clarity": 0.9                           // transcript intelligibility (speech recognition
confidence)
    }
},
"semantic_dna": {
    "catchphrases": ["Let's dive in","As always"], // Recurring phrases or
slogans
    "hook_style": "rhetorical_question",           // How videos start
(question, bold claim, story)
    "narrative_structure": "problem-solution",    // e.g. listicle, story,
tutorial
    "complexity_level": "intermediate",          // 'beginner', 'intermediate',
'expert'
    "topics": ["productivity_tips","devtools"],   // Topic clusters, e.g. via
LDA on transcripts
    "keywords": ["AI","Notion","productivity"],    // High-frequency terms
    "sentiment": {
        "valence": 0.7,                         // [+1: very positive, -1: negative]
        "arousal": 0.5                          // [0-1] emotional intensity
    }
}

```

```

        },
        "writing_style": {
            "first_person": true,           // Uses "I/we" vs impersonal voice
            "humor": 0.3,                  // [0-1] rated humor content
            "expository": 0.5,             // [0-1] more narrative vs factual
            "technicality": 0.6           // [0-1] jargon vs layman terms
        }
    },
    "engagement_dna": {
        "call_to_action_freq": {
            "like": 0.7,                 // Probability of asking to like/subscribe per video
            "comment": 0.2,
            "share": 0.1
        },
        "community_tone": {
            "friendly": true,            // Tone when addressing audience
            "formal": false
        },
        "interaction": {
            "comment_response": 0.3, // Fraction of comments replied
            "user_poll_usage": false
        }
    },
    "behavioral_dna": {
        "upload_schedule": {
            "frequency": "weekly",     // daily/weekly/irregular
            "days_active": ["Tue", "Thu"], // usual days of week
            "time_of_day": "evening"   // e.g. evening, morning
        },
        "video_length": {
            "long_form_avg": 12.5,      // minutes
            "short_form_avg": 1.2       // average short video length (if applicable)
        },
        "platforms": ["YouTube", "Instagram", "LinkedIn"], // Platforms used
        "cross_promotion": {
            "repurpose_rate": 0.4     // How often content is re-used on other
platforms
        }
    },
    "version": 1.0                // Schema version
}

```

Each block in this schema is designed to reflect a slice of the creator's style. For instance, `visual_dna.cut_length_stats` or `color_palette` come directly from video analysis methods <sup>1</sup>.

Similarly, `semantic_dna.topics` and `sentiment` would be populated by NLP on transcripts, leveraging known techniques for topic modeling and sentiment classification [3](#) [2](#). The `audio_dna` fields use pitch and timing metrics (see e.g. speech prosody analysis). The `engagement_dna` section encodes audience-facing habits (like frequency of call-to-actions) gleaned from content and metadata. This JSON is both machine-readable (for programmatic drift-detection and recommendations) and human-readable (comments explain each attribute).

**Updating & Adapting:** In practice, the persona values would be **recomputed regularly** (e.g. rolling average of last N videos) to spot shifts. For example, if `visual_dna.avg_shot_length_sec` increases significantly, the tool could flag “edits are getting slower.” To adapt to different platforms, one could keep the **core persona** and apply transformation rules: e.g. *YouTube* uses full `visual_dna`; *TikTok* might compress videos to 9:16 and increase pacing. The schema even hints at that: the `behavioral_dna.platforms` section lists where and how to repurpose content.

This exhaustive persona profile sets us apart: unlike generic clitters or A/B editors, we capture *why* a creator's content is unique. By codifying their "creative signature" – from average words-per-minute to signature catchphrases – our system can alert them when they drift or even **re-apply their peak style** to new content. The result is a strategic mirror for creators, preserving their identity against algorithmic churn.

4 1

**Sources:** The above schema concepts draw on research in video and language analysis [1](#) [2](#) [3](#) and recent thinking on formalizing digital authorship [4](#). Each category (visual, audio, semantic, etc.) aligns with known AI capabilities (CV, audio DSP, NLP/LLMs), ensuring this persona model is both comprehensive and grounded in current technology.