A Standards-Based Architecture for Advanced Captioning

Using the BBC Subtitle Guidelines as a Compliance Floor for Next-Generation Caption Presentation

Prepared for: Jack Morgan, Matthew Israelsohn, and Claude

Authors: Matthew Israelsohn, and ChatGPT Plus

Source baseline: BBC Subtitle Guidelines, Version 1.2.5, March 2026

Purpose: formal architecture paper for captioning-app and WallSpace planning

This paper treats the BBC guidelines as a standards baseline for readability, timing, synchronisation, positioning, speaker identification, and presentation safety, while allowing structured extension into advanced captioning features such as semantic sound annotation, emotion signalling, expressive rendering, and immersive display modes.

Executive Summary

Core principle: standards-based captioning should be the safety floor, not the innovation ceiling.

- The BBC Subtitle Guidelines must be treated as the baseline requirements source for readable and trustworthy caption presentation, not as a limitation on future functionality.
- The recommended system architecture separates caption input and analysis, canonical caption data, rules enforcement, presentation policy, rendering, and advanced enhancement layers.
- Configurability remains essential, but it should be organised through profile-driven policy logic rather than a flat collection of unrelated switches.
- Advanced features such as emotion cues, semantic sound annotation, reactive typography, immersive placement, and WallSpace-linked visual behaviour must only operate within a protected readability and accessibility floor.
- The first implementation priority should be a formal rules matrix derived from the BBC document, followed by a canonical data model, a rules engine, a presentation policy engine, and regression-tested enhancement layers.

- **System Contract**:
  At no point may rendering or enhancement behaviour reduce caption readability, synchronisation accuracy, or visual safety below the defined compliance floor.

- **Decision Traceability Requirement:**
  All decisions made by the system must be explicitly recorded, structured, and retrievable. This includes rule evaluations, policy resolutions, constraint enforcement, and enhancement adjustments.

- **Decision records must be suitable for:**
  - Human inspection (engineering and review)
  - Claude-based analysis and validation
  - Automated testing and regression tracking

- **Decision Summary:**
  - Build standards-compliant core first
  - Enforce rules via a dedicated engine
  - Enable advanced features only through controlled policy-driven layers

  These architectural documents are intended to be used directly as structured inputs to Claude to ensure implementation and validation align with BBC-based standards.

In practical terms, the project should build a standards-compliant captioning core first, then layer in controlled innovation. The result is a system that can support both formal accessibility requirements and

more ambitious expressive or immersive captioning features without undermining readability, synchronisation, or visual safety.

Contents

Executive Summary 2
1. Introduction 5
2. Core thesis 5
3. Why the BBC guidelines are an excellent baseline 5
4. Scope and relevance 6
5. Architectural model 6
System Invariant 6
5.1 Recommended layer stack 7
5.2 Decision Ownership Model 8
6. Converting guidance into enforceable rule classes 9
Recommended rule classes 9
6.1 Example Rule Definition 9
7. Baseline presentation requirements to enforce 10
8. Configurability and policy control 10
Recommended profile set: 10
Conflict Resolution Priority 11
9. Advanced features beyond the BBC document 11
9.1 Feature Acceptance Criteria 11
10. Compliance floor and enhancement ceiling 11
10.1 Failure Modes and Safeguards 12
Caption too fast 12
→ Risk: unreadable 12
→ Detection: WPM exceeds threshold 12
→ Mitigation: extend duration or trigger edit suggestion 12
11. Validation and Scoring Model 12
12. Decision Logging and Traceability 13
13. Automated and manual testing strategy 14
14. Claude Integration Interface 14

Constraint Enforcement Requirement 15
15. Recommended internal document set 16
16. Practical implementation priorities 16
Recommended implementation order 16
17. Recommended design decisions 16
18. Conclusion 17
Appendix A. Operational Summary 17
Appendix B. Terminology 17
Core Concepts 17
Constraint 17
Compliance Floor 18
Enhancement Layer 18
Canonical Caption Data Model 18
Standards Rules Engine 18
Presentation Policy Engine 18
Renderer 18
Input Engines 18
Validation and Telemetry 19
Behavioural Concepts 19
Profile (Presentation Profile) 19
Rule Class 19
Readable Caption State 19
Segmentation 19
Reflow Risk 19
Safe Region (Safe Area) 19
Speaker Attribution 19
System Behaviour 20
Decision Ownership Model 20
Failure Mode 20
Fallback Behaviour 20
Enhancement Constraint 20

1. Introduction

The BBC Subtitle Guidelines are a particularly strong foundation for this project because they are written not only for subtitle authors and editors, but also for developers of subtitle authoring, validating, converting, and presentation tools. The document therefore functions as both an editorial guide and a practical systems baseline.

The captioning app and WallSpace are not conventional subtitle tools. They aim to coordinate multiple engines, including speech recognition, speaker handling, timing logic, visual layout, sound labelling, and potentially emotional or immersive behaviour. That requires a formal architecture rather than an accumulation of isolated features.

The BBC guidelines are highly relevant because they explicitly frame subtitling as the negotiation of competing constraints: fidelity to speech, readability, timing, synchronisation, visual protection, shot boundaries, and user comprehension. Those same tensions are central to any next-generation captioning system.

Non-Goals

- This system is not limited to traditional subtitle rendering.
- This system does not prioritise visual effects over readability.
- This system does not rely on uncontrolled automatic layout.
- This system does not treat all features as equally enabled.

2. Core thesis

The core thesis of this paper is simple: standards-based captioning should define the minimum safe and readable behaviour of the system, while advanced features should be implemented only as controlled extensions above that baseline.

This approach allows the project to remain both ambitious and trustworthy. It avoids two opposite failures: a purely traditional subtitle engine that cannot evolve, and an over-expressive system that quietly damages readability or accessibility.

The system should therefore make decisions in the following order: readability first, compliance second, enhancement third, and expression fourth.

3. Why the BBC guidelines are an excellent baseline

The BBC document is unusually valuable because it spans text editing, line breaks, timing, synchronisation, shot matching, speaker identification, colour usage, typography, positioning, difficult speech, intonation and emotion, music, sound effects, live subtitling, and delivery formats.

It also contains guidance aimed specifically at implementers. That makes it suitable for conversion into a machine-checkable rules library, rather than being treated as a passive style reference.

Finally, the document already acknowledges that subtitling is contextual and cannot be reduced to rigid universal rules. That makes it compatible with a policy-driven architecture in which different content types and presentation modes can still share one formal baseline.

4. Scope and relevance

Not every part of the BBC document is equally relevant to the runtime behaviour of the app. The most important sections for architecture are the presentation-oriented areas: text editing, line breaks, timing, synchronisation, matching shots, identifying speakers, colours, typography, positioning, intonation and emotion, difficult speech, music, sound effects, cumulative subtitles, and live subtitling.

The technical format sections remain useful wherever the system needs to export, validate, or interoperate with standard subtitle formats, but some legacy Teletext-specific delivery constraints are less central to the immediate runtime architecture.

Even where a legacy format is not directly used, its historical constraints often explain why certain timing, line length, or sizing recommendations exist. Those parts should therefore be treated as contextual background rather than discarded outright.

5. Architectural model

The system should not be designed as a single path from speech input to visible captions. It should be designed as a layered system: input and analysis engines, canonical caption data model, standards rules engine, presentation policy engine, renderer, and optional enhancement layers.

Input and analysis engines include, but are not limited to, speech recognition, OCR, imported subtitle files, speaker diarisation, sound-event detection, shot detection, confidence scoring, and emotion or prosody inference. These engines should produce structured evidence, not final display behaviour.

All upstream results should be normalised into a canonical caption data model. At minimum, that model must include timing, text, speaker identity, source type, caption type, confidence, line-break candidates, synchronisation context, shot and scene references, style suggestions, and enhancement eligibility metadata.

A standards rules engine should then validate, score, and where safe repair caption behaviour against baseline requirements. A presentation policy engine should use profiles and context to determine the final rendering strategy. The renderer produces the final output. Enhancement layers may apply controlled modifications to rendering behaviour, subject to validation against rules and policy constraints.

Minimum Required Fields (must be present for all captions):

- start_time
- end_time
- text
- speaker_id (or explicit unknown)
- caption_type (speech / sound / other)

Optional fields may be omitted, but required fields must always be present to ensure baseline validation.

System Invariant

The following invariant applies across all system layers:

All caption processing, rendering, and enhancement operations must preserve the compliance floor at all times.

All decisions affecting caption behaviour must be recorded and traceable.

No component may override or bypass these constraints.

5.1 Recommended layer stack

Category	Description
Layer 1	Input and analysis engines
Layer 2	Canonical caption data model
Layer 3	Standards rules engine
Layer 4	Presentation policy engine
Layer 5	Renderer and enhancement layers

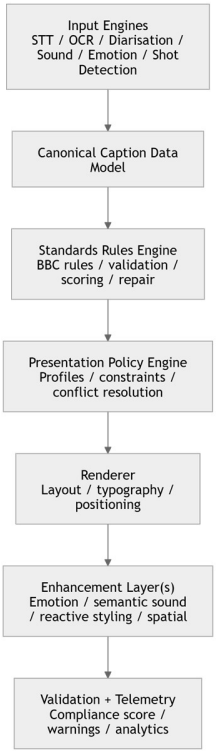

Figure 1 - Captioning System Architecture (Baseline + Enhancement Flow)
End-to-end data flow showing how multiple input engines are normalised into a canonical caption model, validated against standards, resolved through policy, rendered for display, and optionally extended via controlled enhancement layers, with validation and telemetry ensuring ongoing compliance and system integrity.

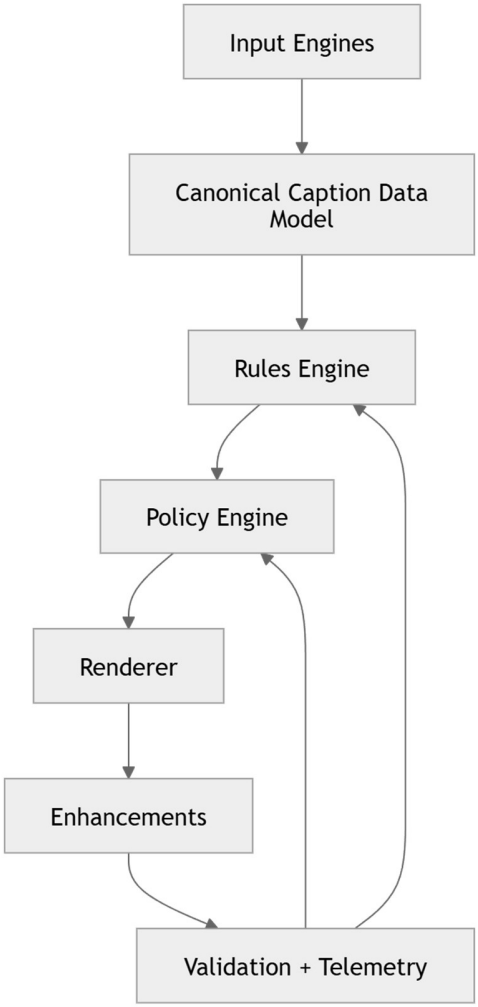

Figure 2 — Feedback-Driven Captioning Pipeline (Closed-Loop Control)
Closed-loop architecture showing how validation and telemetry feed back into the rules and policy engines, enabling continuous adjustment of caption behaviour, enforcement of compliance constraints, and dynamic optimisation of rendering and enhancement decisions.

5.2 Decision Ownership Model

- **Input Engines**: generate structured data only (no UI or rendering decisions)
- **Rules Engine**: validates data and enforces constraints
- **Policy Engine**: determines rendering behaviour based on profiles and context
- **Renderer**: executes layout and display decisions
- **Enhancement Layer**: proposes controlled modifications (must not enforce behaviour)

All enhancement proposals must be validated against rules and policy constraints before application.

All components must emit structured decision logs for every action affecting caption behaviour.

6. Converting guidance into enforceable rule classes

The BBC document should be converted into a formal rules matrix rather than left as a long prose reference. Each rule must receive a stable identifier, source section, summary, category, severity, machine-checkability rating, and suggested validation metric.

The rules must be organised into at least four classes. Hard accessibility rules define the compliance floor. Preferred presentation rules define strong defaults with context-sensitive exceptions. Controlled expressive rules cover the restrained signalling already recognised by the guidelines. Project-specific advanced rules govern features beyond the BBC document, such as reactive typography, semantic sound layers, or immersive spatial placement.

This structure keeps innovation disciplined. New behaviours can be introduced without weakening the baseline logic that governs legibility, synchronisation, picture protection, and speaker clarity.

Recommended rule classes

Category	Description
Class A	Hard accessibility rules
Class B	Preferred presentation rules
Class C	Controlled expressive rules
Class D	Project-specific advanced rules

6.1 Example Rule Definition

Rule ID: BBC-TIM-001
Category: Timing
Class: A (Hard Accessibility)

Description:
Subtitle must allow sufficient reading time based on word count.

Metric:
WPM ≤ 180 OR duration ≥ 0.33s per word

Failure Condition:
Computed reading speed exceeds threshold

Auto-fix:
Extend duration if possible, otherwise trigger edit suggestion

Severity:
High

7. Baseline presentation requirements to enforce

The BBC guidance strongly prefers verbatim text when time allows, rejects condescending simplification, values retention of lip-readable words, and expects the speaker's style and names to be preserved where possible. The default editing strategy should therefore be minimal and meaning-preserving, not aggressively compressive.

Line breaking should be deliberate rather than automatic. The BBC explicitly warns against splitting tightly bound grammatical units and says implementers should avoid relying on automatic line wrapping. The project should therefore use explicit segmentation logic, not browser-like opportunistic wrapping.

Timing rules should reflect both the recommended baseline of roughly 160 to 180 words per minute and the contextual reasons to vary from it, such as unfamiliar words, several speakers, labels, dense visual information, shot changes, and slow speech. Synchronisation checks should ensure close alignment with speech onset, visible presence during lip movement, and bounded lag.

Speaker identification, colour behaviour, typography, sizing, and positioning must therefore be treated as core validation targets. In particular, the system should preserve safe regions, avoid obscuring mouths and important graphics, and never permit final presentation size to exceed authored size.

8. Configurability and policy control

Configurability is essential in a captioning system that may operate across broadcast-like, live, enhanced, immersive, and experimental modes. However, configurability should be managed through a presentation policy layer rather than a flat collection of unrelated toggles.

The recommended approach is profile-driven. At minimum, the system should support: Compliance or Broadcast-safe, Accessible Enhanced, Live Fast or Low Latency, Immersive or Spatial, and Experimental or Expressive profiles. Each profile should enable or restrict classes of behaviour according to the required trust level and presentation goal.

This prevents contradictory states. For example, aggressive motion, large-scale dynamic emphasis, strict mouth-avoidance, and compact mobile layout may not all be safely compatible. A policy engine can resolve those tensions consistently.

Recommended profile set:

Category	Description
Compliance / Broadcast-safe	Conservative BBC-aligned behaviour
Accessible Enhanced	Low-risk semantic additions on top of the baseline
Live Fast / Low Latency	Real-time readability under live constraints
Immersive / Spatial	Richer placement within motion and safety limits
Experimental / Expressive	R&D mode with baseline guardrails still intact

Conflict Resolution Priority

When multiple constraints conflict, the system must resolve them using the following strict priority order:

1. Readability and accessibility
2. Synchronisation and timing accuracy
3. Visual safety (non-obstruction)
4. Speaker clarity
5. Enhancement behaviour

Lower-priority behaviours must yield to higher-priority constraints without exception.

9. Advanced features beyond the BBC document

The BBC document does not attempt to fully specify features such as emotion engines, continuous prosody-linked modulation, volume-reactive size changes, immersive three-dimensional placement, semantic sound overlays, caption-driven graphics, or WallSpace-linked visual control.

That absence should not be treated as a prohibition. Instead, such features should be admitted only when they solve a real user problem, have an intelligible fallback, can be reduced or disabled, and do not silently reduce comprehension.

For example, volume-reactive size change may convey emphasis or intensity, but it can also destabilise text, create reflow, and obscure image details. A controlled implementation would therefore clamp the size range, smooth transitions, prevent mid-caption reflow, and disable the effect in stricter profiles.

Advanced features must degrade gracefully when required data is incomplete, unreliable, or unavailable.

9.1 Feature Acceptance Criteria

Every advanced feature must satisfy all of the following conditions:

- Improves comprehension OR adds meaningful context
- Does not reduce readability
- Has bounded behaviour (limits defined)
- Has a safe fallback
- Can be disabled or scaled
- Does not conflict with baseline rules
- Works under low-confidence conditions

10. Compliance floor and enhancement ceiling

A formal safety statement must exist in the project documentation: all rendering modes must preserve a minimum readability and accessibility baseline defined by the active compliance profile, and enhancements may operate only within that boundary.

At minimum, the compliance floor should monitor reading speed, visible duration, synchronisation lag, line count, line length, contrast, reflow risk, safe-area adherence, speaker clarity, mouth and graphic obstruction risk, and positional stability.

This enforces a clear separation between baseline caption safety and optional enhancement behaviour.

10.1 Failure Modes and Safeguards

Each failure mode must be handled using the following pattern: Condition → Risk → Detection → Mitigation

Example:

Caption too fast

→ Risk: unreadable

→ Detection: WPM exceeds threshold

→ Mitigation: extend duration or trigger edit suggestion

11. Validation and Scoring Model

The system should not rely on binary pass or fail checks alone. Each caption or caption sequence should be scored across multiple dimensions, such as text fidelity, segmentation quality, timing adequacy, synchronisation quality, speaker clarity, visual obstruction risk, typography safety, layout stability, and enhancement safety.

Validation output should be human-usable. A caption should be classifiable as compliant, compliant with warnings, partially compliant, non-compliant but recoverable, or non-compliant and unsafe. That gives engineers and reviewers a concrete basis for comparison.

Decision transparency must be part of validation output.

Each validation result should include:

- the decision path that led to the outcome
- the rules and constraints applied
- any trade-offs made during conflict resolution

Example Flow:

Input:

- Rapid speech at 220 WPM

Processing:

- Rules Engine detects timing violation (WPM exceeds threshold)
- Policy Engine prioritises readability over verbatim fidelity
- Renderer applies segmentation and timing adjustment

Enhancement:

- Emotion layer proposes emphasis styling
- Policy Engine reduces intensity to preserve readability constraints

Output:

- Readable, compliant caption with reduced expressive styling

12. Decision Logging and Traceability

All system decisions must be logged in a structured and consistent format.

Testing must verify not only outcomes but also decision correctness.

Test cases must confirm:

- correct rules were applied
- correct conflict resolution order was followed
- decision logs accurately reflect system behaviour

A "decision" includes any operation where the system:

- applies or evaluates a rule
- resolves a conflict between constraints
- modifies caption content or timing
- applies or suppresses an enhancement
- selects a presentation policy outcome

Each decision record must include:

- Decision type (rule / policy / enhancement / rendering)
- Input state (relevant caption data and context)
- Applied rules or constraints (with rule IDs)
- Decision outcome
- Reasoning (human-readable explanation)
- Confidence level (if probabilistic)
- Timestamp and processing stage

Decision logs must be:

- Machine-readable (for Claude and automation)
- Human-readable (for debugging and review)
- Linked to specific captions or sequences

The system must allow reconstruction of the full decision chain for any caption.

13. Automated and manual testing strategy

Three levels of testing are recommended. Rules tests verify that the standards engine correctly detects issues such as excessive word rate, destructive line breaks, sync drift, unsafe overhang, shot straddling, or speaker ambiguity.

Rendering tests verify that the interface actually respects the chosen policy: safe zones, font scaling, overlay avoidance, stable alignment, and profile-dependent feature gating.

Perception tests are especially important for advanced features. Expressive or immersive behaviours should be evaluated for comprehension, comfort, fatigue, trust, and user preference rather than assumed to be beneficial simply because they are richer.

All tests must include failure-case scenarios to verify correct fallback behaviour and must be documented.

14. Claude Integration Interface

Claude is used for the following tasks:

1. Rule extraction
2. Rule-to-system mapping
3. Test case generation
4. Compliance scoring
5. Gap analysis

Example prompt pattern:

Input:
- BBC rules matrix
- Current caption data
- Active policy profile

Task:
- Evaluate compliance
- Identify violations
- Suggest fixes
- Output structured report

Expected Output Format:

- Compliance status (per caption)

- Rule violations (with rule IDs)

- Severity classification

- Suggested fixes

- Confidence level (if applicable)

Claude must be able to consume and reason over decision logs.

Claude tasks should include:

- auditing decision chains
- identifying incorrect rule applications
- detecting inconsistent policy behaviour
- validating enhancement safety decisions

Constraint Enforcement Requirement

Claude outputs must respect the defined conflict resolution priority and compliance floor. Any suggested changes that violate these constraints must be rejected or flagged.

14.1 Claude as Execution and Validation Layer

Claude is not used as a general-purpose assistant within this system. It is treated as a constrained reasoning and validation engine operating strictly within the defined architecture.

All documents in this system (Rules Matrix, Canonical Caption Data Model, Policy Engine, Renderer, Enhancement Layer, and Decision Log Specification) are intended to be provided to Claude as authoritative inputs.

Claude must:

- Interpret BBC-derived rules via the Rules Matrix
- Evaluate caption data against those rules
- Apply policy-driven decision logic
- Generate structured outputs aligned with the Decision Log specification
- Identify violations, gaps, and inconsistencies
- Propose fixes that remain within system constraints

Claude must not:

- Invent new rules or behaviours outside the supplied specifications
- Override defined constraint priorities
- Introduce presentation behaviour not governed by the Policy Engine

All Claude outputs must be:

- Structured and machine-readable
- Traceable to rules and constraints
- Compatible with the Decision Log system

This ensures that Claude acts as a controlled execution layer for caption validation, transformation, and system reasoning, rather than an unconstrained generative component.

15. Recommended internal document set

The project should maintain a Caption Rules Matrix, a Canonical Caption Data Model Specification, a Presentation Policy Specification, an Enhancement Safety Specification, a Compliance Gap Report, a Test Plan, and an Exception Register.

Decision Log Schema

A formal definition of how system decisions are recorded, structured, and stored.

Together, these documents create a durable architecture spine. They also ensure that future features are attached to an explicit framework rather than quietly altering baseline behaviour.

16. Practical implementation priorities

The recommended build order is: first, extract and formalise the BBC-based rules matrix. Second, ensure the canonical caption data model contains enough information to validate against those rules. Third, implement a standards rules engine. Fourth, introduce a policy engine that controls all configurability. Fifth, harden the baseline renderer. Sixth, add advanced features only through guarded extension points. Seventh, regression-test every enhancement against the compliance floor.

This order matters. It ensures that innovation builds on a stable and auditable core instead of being used to compensate for baseline uncertainty.

Recommended implementation order

Category	Description
Phase 1	Extract and formalise BBC-based rules
Phase 2	Build and validate the canonical caption model
Phase 3	Implement the rules engine
Phase 4	Implement the presentation policy engine
Phase 5	Harden the baseline renderer
Phase 6	Add guarded enhancement layers
Phase 7	Regression-test every enhancement

17. Recommended design decisions

Automatic line wrapping should not be trusted. Deliberate segmentation and layout control are preferable.

Final presentation size should never exceed authored size. User configurability should work within that ceiling rather than above it.

Enhancements should not override speaker identification or picture protection rules. Speaker clarity and visual safety remain foundational.

Live mode should not be treated as prepared mode with lower latency. It has materially different constraints and should have its own presentation policy profile.

18. Conclusion

The BBC Subtitle Guidelines provide a current, wide-ranging, and implementation-friendly baseline for the design of an advanced captioning system. They are especially valuable because they are written for both editorial practitioners and software developers.

The correct architectural response is therefore not to reduce the system to a conventional subtitle engine, but to separate it into a standards-compliant captioning core and a controlled innovation layer. That allows the project to pursue advanced caption intelligence without losing trust, clarity, or accessibility.

The enduring design principle is therefore: use the BBC guidelines as the compliance floor, govern all configurability through a policy engine, and allow advanced behaviour only when readability, synchronisation, and visual safety remain protected.

This architecture is designed to be both enforceable and extensible, ensuring that future innovation can be introduced without compromising established accessibility guarantees.

The architecture is designed to be directly implementable, testable, and extensible within both real-time and pre-authored captioning systems.

Appendix A. Operational Summary

Decision Order (mandatory):

1. Ensure readability and accessibility
2. Enforce compliance
3. Apply enhancement constraints
4. Apply expressive behaviour (if permitted)

Input engines -> canonical caption model -> rules engine -> policy engine -> renderer -> enhancement layers -> validation and telemetry.

BBC guidance defines the minimum acceptable behaviour.

Project-specific innovation may extend beyond it, but never beneath it.

Appendix B. Terminology

Core Concepts

Constraint

A rule or condition enforced by the system that limits behaviour to preserve readability, synchronisation, accessibility, or visual safety. Constraints are derived from rules, applied by the rules engine, and enforced through the policy engine.

Compliance Floor

The minimum level of readability, accessibility, timing accuracy, and visual safety that must be preserved in all caption rendering modes. Derived primarily from standards such as the BBC Subtitle Guidelines. No feature or enhancement may reduce behaviour below this baseline.

Enhancement Layer

A controlled set of optional features that extend baseline caption behaviour, such as emotion signalling, semantic sound annotation, reactive typography, or spatial positioning. Enhancements must operate within constraints defined by the compliance floor and policy engine.

Canonical Caption Data Model

A structured internal representation of caption data that normalises inputs from all upstream systems (e.g. STT, OCR, imported files). It contains all information required for validation, policy decisions, and rendering, including timing, text, speaker identity, metadata, and analysis outputs.

Decision Log

A structured record of a system decision, including inputs, applied rules, constraints, outcomes, and reasoning. Used for debugging, validation, auditing, and AI-assisted analysis.

Standards Rules Engine

A system component that enforces captioning standards by validating, scoring, and optionally repairing caption data. It converts guideline-based rules (e.g. BBC standards) into machine-checkable logic and produces structured compliance outputs.

Presentation Policy Engine

A decision layer that determines how captions should be rendered based on context, active profile, constraints, and user configuration. It resolves conflicts between readability, compliance, and enhancement behaviour.

Renderer

The component responsible for producing the final visual output of captions, including layout, typography, positioning, timing execution, and display behaviour across different platforms or environments.

Input Engines

Independent systems that generate caption-related data, such as speech-to-text, OCR, speaker diarisation, sound detection, emotion inference, and shot detection. These engines provide structured inputs but do not control final presentation.

Validation and Telemetry

A system layer that monitors caption performance, including compliance scoring, rule violations, rendering behaviour, and user interaction data. Used for debugging, optimisation, and regression testing.

Behavioural Concepts

Profile (Presentation Profile)

A predefined configuration that controls how captions are rendered, including which rules, constraints, and enhancements are active. Examples include Compliance, Enhanced, Live, Immersive, and Experimental modes.

Rule Class

A categorisation of captioning rules based on their importance and flexibility:

- Class A: Hard accessibility rules (must not be violated)
- Class B: Preferred presentation rules (strong defaults)
- Class C: Controlled expressive rules (limited stylistic behaviour)
- Class D: Advanced project-specific rules (experimental features)

Readable Caption State

A condition in which a caption can be comfortably read and understood, considering timing, segmentation, contrast, positioning, and stability.

Segmentation

The process of dividing speech or text into caption units and lines based on linguistic, timing, and visual constraints. Must prioritise comprehension over mechanical splitting.

Reflow Risk

The likelihood that text layout changes (e.g. due to scaling or dynamic styling) will alter line breaks or positioning in a way that reduces readability or stability.

Safe Region (Safe Area)

The portion of the screen where captions can be displayed without obscuring important visual content such as faces, graphics, or UI overlays.

Speaker Attribution

The process of identifying and distinguishing speakers within captions using colour, labels, positioning, or metadata.

System Behaviour

Decision Ownership Model

Defines responsibility boundaries within the system:

- Input Engines → generate data
- Rules Engine → enforce constraints
- Policy Engine → decide behaviour
- Renderer → execute display
- Enhancement Layer → propose controlled modifications

Failure Mode

A condition where caption behaviour degrades (e.g. unreadable timing, layout instability, loss of speaker clarity). The system must detect and correct or mitigate these conditions.

Fallback Behaviour

A safe alternative rendering mode used when ideal behaviour cannot be achieved, ensuring captions remain readable and compliant.

Enhancement Constraint

A rule that limits advanced feature behaviour to prevent degradation of readability, synchronisation, or accessibility.