Formatting Instructions for ICLR 2027 Conference Submissions

Anonymous authors
Paper under double-blind review

Abstract

Automatic video chaptering organizes long instructional videos into reusable directories for efficient content access. Current models typically learn from annotated directories through supervised fine-tuning (SFT) or reference-based optimization. However, a reference directory represents only one plausible organization of a video, and optimizing agreement with it can favor annotation-specific predictions over alternative but equally valid chapter structures. More importantly, higher reference agreement does not necessarily imply greater downstream utility: a directory closer to the ground truth may not be more effective for helping users locate the content they need. We introduce ChapterNav, a framework for evaluating and optimizing video chaptering through verifiable navigation utility. Since chaptering admits multiple valid outputs and lacks a directly verifiable target sequence, ChapterNav instead measures whether a generated directory enables a fixed navigator to locate content satisfying a given information need. Given a query and chapter titles, the navigator selects chapters, while a verifier checks whether the retrieved intervals contain the target content within a prescribed viewing budget. We build navigation benchmarks on two datasets and introduce Successful Navigation Gain (SNG) to quantify navigation utility. The resulting feedback, combined with temporal alignment rewards, is further used for reinforcement learning with verifiable rewards (RLVR). In a paired human study on 50 videos, SNG agrees with human navigation preferences on 76% of pairs, compared with 68% for reference-based evaluation. RLVR achieves a relative improvement of 3.8% on the SNG metric compared with SFT, and a relative improvement of 5.2% compared with reference-reward. These results demonstrate the effectiveness of evaluating and learning video chaptering through downstream content-access utility.

1 Introduction

Long-form videos, such as lectures, tutorials, and meeting recordings, contain rich information, yet users often seek only the portions relevant to a particular goal. Supporting efficient access therefore requires organizing video content into temporally grounded and semantically meaningful units. Dense video captioning Krishna et al. (2017); Wang et al. (2021a); Kim et al. (2024; 2025); Baek et al. (2026) follows a localize-and-describe paradigm, identifying events in untrimmed videos and describing them, whereas video chaptering Yang et al. (2023a); Ventura et al. (2025) emphasizes higher-level organization by partitioning a video into consecutive chapters with titles or descriptions for browsing and navigation. Unlike query-conditioned temporal localization Li et al. (2025); Gao & Xu (2021); Moon et al. (2023), chapter generation produces a directory without knowing the particular information need it will subsequently serve. An effective chapter directory should therefore serve as a reusable navigation structure, providing informative entry points that help users locate complete target content while minimizing unnecessary viewing.

Recent progress in video chaptering has been driven by stronger large language models, and increasingly sophisticated structural learning objectives. VidChapters-7M Yang et al. (2023a) provides large-scale chapter annotations, while Chapter-Llama Ventura et al. (2025) leverages long-context language models to jointly generate chapter boundaries and titles from timestamped transcripts and visual descriptions. More recent approaches further improve hierarchical structure modeling, temporal alignment, and cross-chapter dependency reasoning Pu et al. (2025); Duan et al. (2026). Despite these advances, both training and evaluation remain largely anchored to agreement with reference

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

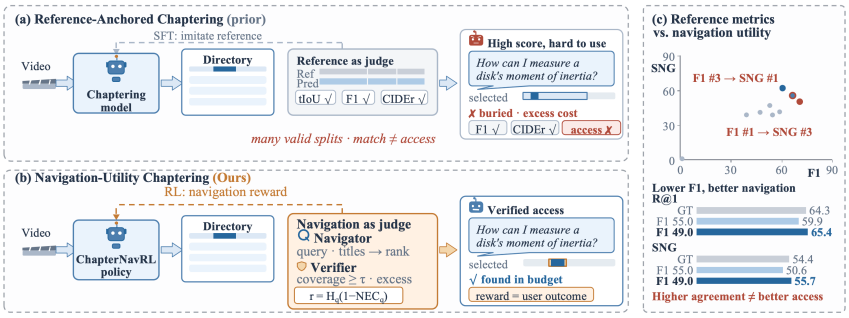

Figure 1: (a) Prior chaptering systems are trained and scored by agreement with a single reference partition, so a high-scoring directory can still bury the content a query needs. (b) ChapterNavRL instead rewards verifiable navigation utility: a frozen navigator ranks chapter titles against the query, and a verifier checks target coverage under a viewing budget, so the reward coincides with the user's retrieval outcome. (c) Reference agreement broadly tracks utility across model scales, yet it misranks competitive systems, a less reference-similar directory can even outperform the reference itself.

chapter annotations, treating reference consistency as a proxy for chapter quality. This raises a fundamental question: does a directory that better matches a reference organization necessarily provide more effective access to the underlying video?

Reference-based metrics and navigation utility capture different properties of a chapter directory. Existing metrics primarily measure agreement with annotated boundaries, intervals, or textual descriptions, whereas successful navigation additionally requires that chapter titles guide users to the appropriate interval, that the selected interval contains the required content, and that this content can be accessed within a reasonable viewing budget. Consequently, a prediction can match a reference well without necessarily providing better content access, while an alternative chapter organization may support effective navigation despite deviating from the annotation. Our navigation-based evaluation confirms this mismatch, revealing that higher reference-based scores do not consistently correspond to better navigation outcomes [Fig. 1]. This discrepancy also affects learning: optimizing models to better reproduce a reference directory does not necessarily reward the chapter structures that are most useful for downstream access.

Although chapter organization is inherently open-ended, the navigation outcomes enabled by a generated directory can be verified under a well-defined access task. Instead of judging whether a directory matches a particular reference partition, we ask whether it allows a user to locate the content required by a specified information need. This shifts supervision from reproducing one annotated structure to verifying what the generated structure enables downstream. By grounding success in content access rather than reference similarity, the resulting feedback can accommodate multiple valid chapter organizations while still providing an objective learning signal.

We introduce ChapterNav, a framework for learning video chaptering through verifiable navigation feedback. Given a video representation, the chaptering model first generates a directory without access to the downstream queries used for evaluation. A separate, fixed navigator then selects chapters using only a query and the generated chapter titles, while a verifier checks whether the selected intervals cover the required content and supporting evidence within predefined viewing budgets. These outcomes are aggregated into a directory-level navigation reward and used to optimize the chaptering model through reinforcement learning with verifiable rewards (RLVR) Lambert et al. (2024). Importantly, queries supervise chapter generation only through the reward, preserving the goal of learning a reusable directory rather than producing query-specific temporal predictions.

To support both evaluation and learning, we construct navigation benchmarks on two datasets, pairing videos with information-seeking queries and grounded target evidence defined independently of any generated directory. We introduce Successful Navigation Gain (SNG), which measures whether a directory enables successful content access under controlled viewing budgets, and evaluate both existing chaptering systems and utility-driven learning under the same protocol. In a paired human study on 50 videos, SNG agrees with human navigation preferences on 76% of comparisons, versus 68% for reference-based evaluation. RLVR with this feedback further improves SNG over both SFT and reference-reward RL. Together, these results show that navigation outcomes provide both a more direct measure of downstream chapter utility and an effective learning signal for chapter generation.

Our contributions are threefold: (1) We establish and empirically characterize a mismatch between reference-based chapter quality and downstream navigation utility, showing that agreement with a reference directory is an imperfect proxy for effective content access. (2) We introduce ChapterNav, which converts open-ended chapter generation into verifiable navigation outcomes and uses these outcomes as rewards for reinforcement learning, while preserving query-agnostic directory generation. (3) We construct navigation benchmarks and a unified evaluation protocol based on grounded information needs, enabling chaptering systems to be assessed by the content access they support rather than solely by agreement with a reference partition.

2 Related Work

Video Chaptering. Dense video captioning aims to localize and describe events in untrimmed videos Krishna et al. (2017), with models such as Vid2Seq Yang et al. (2023b) jointly generating temporal boundaries and descriptions. Video chaptering further organizes an entire video into contiguous, non-overlapping semantic units for high-level browsing and navigation. Recent benchmarks and models, including VidChapters-7M Yang et al. (2023a) and Chapter-Llama Ventura et al. (2025), explore large-scale chapter generation with long-context language models, while subsequent studies investigate hierarchical structuring and cross-segment temporal reasoning Li et al. (2026); Pu et al. (2025); Duan et al. (2026). These works mainly focus on generating chapter structures and evaluating their agreement with annotations, whereas our work studies whether generated chapters effectively support downstream content access.

Chapter and Caption Evaluation. Existing evaluation protocols mainly compare generated chapters or captions with human references. Temporal metrics such as tIoU evaluate boundary accuracy and interval overlap, while language metrics such as BLEU Papineni et al. (2002), METEOR Banerjee & Lavie (2005), CIDEr Vedantam et al. (2015), and SODA Fujita et al. (2020) measure semantic similarity and temporal alignment with reference descriptions. Although these metrics quantify structural and semantic consistency, they do not directly assess whether generated chapters help users locate desired information in long videos. Recent studies evaluate generated descriptions through downstream tasks such as question answering Qiao et al. (2024); Yang et al. (2025); Lin et al. (2026); however, chapter directories differ in that they serve as navigation structures rather than evidence-bearing descriptions. Therefore, ChapterNav evaluates whether generated chapters can guide users to relevant video segments under a constrained browsing budget.

Reinforcement Learning with Verifiable Rewards. Reinforcement learning with verifiable rewards (RLVR) optimizes models using automatically checkable task outcomes Lambert et al. (2024), and has been particularly effective in domains such as mathematical reasoning and code generation Shao et al. (2024); DeepSeek-AI et al. (2025); DeepSeek-AI (2024). Open-ended generation is more challenging because multiple outputs may be valid and sequence-level quality is difficult to verify directly. Recent approaches address this issue by defining rewards through downstream tasks enabled by generated content Xing et al. (2026); Yang et al. (2026). Following this direction, ChapterNav verifies whether a generated directory enables complete and efficient access to target video content and uses the resulting navigation outcomes as rewards for chapter generation, without requiring agreement with a single reference partition.

3 Methodology

3.1 Task Formulation

We represent a video V of duration T as time-aligned text $X = \mathcal{T}(V)$, which can include speech transcripts, OCR text, or visual descriptions. A chaptering model maps this representation to an ordered directory:

$$C = f_\theta(X) = \{(s_j, e_j, t_j)\}_{j=1}^{M}, \tag{1}$$

where s_j, e_j, and t_j denote each chapter's start time, end time, and title. The chapters form a continuous partition with $s_1 = 0$, $e_M = T$, and $e_j = s_{j+1}$, and their titles describe high-level content.

Navigation evaluates a fixed directory against subsequent queries. Each query q has an annotated target episode $E_q = [a_q, b_q]$, and the directory is generated without access to q by the chaptering model. A navigator ranks chapters using only the query and chapter titles; target-episode annotations

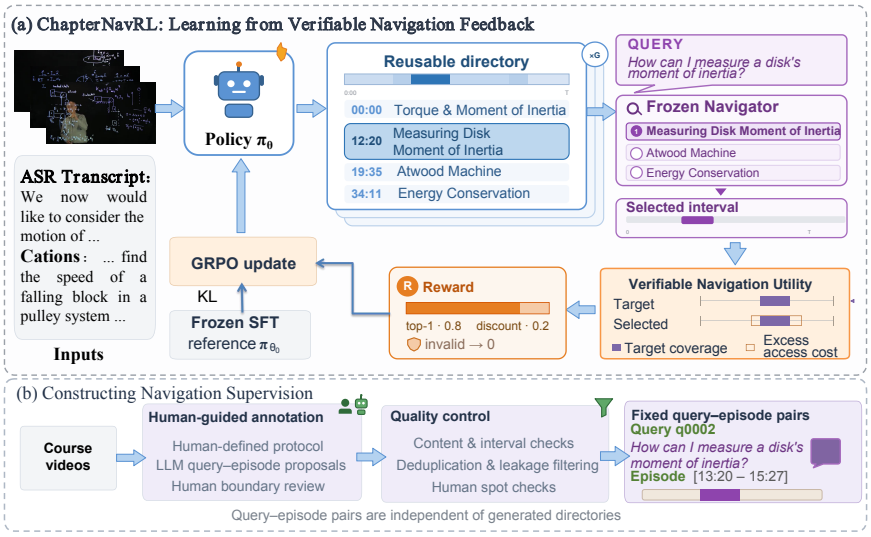

Figure 2: **Overview of ChapterNav.**(a) A chaptering policy generates complete, query-agnostic directories. For each query, a frozen navigator ranks chapter titles and selects an interval; a verifier measures target coverage and excess access cost. The resulting rewards guide GRPO, with KL regularization to a frozen SFT reference.(b) Human-guided annotation and quality control produce fixed query-episode pairs independently of generated directories. Training pairs provide rewards, while held-out pairs support evaluation.

verify navigation outcomes and, on the training split, provide learning feedback. Unlike query-conditioned temporal localization, this setting evaluates whether a reusable directory enables users to access relevant video content through its chapter entries.

3.2 Constructing Navigation Supervision

ChapterNav adds navigation annotations to two collections: AVLecture Gupta et al. (2023), an existing chaptering dataset, and KZKT, public Chinese K-12 course videos gathered from internet platforms. As shown in Fig. 3.1 (b), each instance pairs a video with a natural-language query and a target episode containing the content needed to address it, rather than using original chapter annotations as queries. Sources, collection procedures, statistics, and usage considerations appear in the appendix.

Human annotators and LLM agents jointly construct the query-episode pairs. Human annotators first define the protocol, specifying what constitutes a natural navigation query and how its target episode should be delimited. Following this protocol, LLM agents inspect video content, discover candidate episodes, formulate corresponding queries, and refine their temporal boundaries. During annotation, humans examine query-episode correspondence and manually check the proposed boundaries. The queries describe needs that can be satisfied by visiting localized video content, independently of directories produced by evaluated models. The resulting pairs are fixed and shared across generators, allowing their directories to be compared under the same navigation needs.

Quality control combines content checks, filtering, and human spot checks. Agent-based checks assess whether queries express clear, self-contained needs, whether episodes provide sufficient content to address them, and whether boundaries exclude unrelated context. Programmatic filters remove invalid intervals, near-duplicate queries, and direct textual leakage from source chapter metadata. After construction, humans inspect sampled instances to identify systematic annotation problems and assess the collection's suitability for navigation evaluation. The appendix provides annotation instructions, agent prompts, filtering rules, inspection procedures, and examples.

3.3 Navigation Evaluation

We evaluate navigation by ranking chapter titles and verifying target coverage. A frozen navigator model such as MiniLMv2 Wang et al. (2021b) or Qwen3-4B-Embdeding Zhang et al. (2025) ranks

titles by query similarity, yielding chapter-index permutation ρ_q. Let $I_j = [s_j, e_j]$ and $\widehat{I}_q = I_{\rho_q(1)}$ denote chapter intervals and the first-ranked interval, respectively. Temporal recall measures coverage:

$$\text{tR}_{qj} = |E_q \cap I_j|/|E_q|, \tag{2}$$

where $|\cdot|$ measures duration. A chapter qualifies when $\text{tR}_{qj} \geq \tau$, with $\tau = 0.9$ fixed throughout evaluation. This criterion provides a computable outcome from annotated intervals rather than an open-ended LLM judgment.

We distinguish chapter availability from successful title routing using

$$A_q = \mathbb{1}\left[\max_j \text{tR}_{qj} \geq \tau\right], \qquad H_q = \mathbb{1}\left[\text{tR}_{q,\rho_q(1)} \geq \tau\right]. \tag{3}$$

Across N queries, the corresponding metrics are

$$\text{BestTR} = \sum_q A_q/N, \quad \text{R@1} = \sum_q H_q/N, \quad \text{CondTR} = \sum_q H_q/\sum_q A_q. \tag{4}$$

BestTR measures the fraction of queries with at least one qualifying chapter, R@1 measures first-choice success, and CondTR measures successful routing conditional on availability.

We additionally measure access efficiency because a whole-video chapter satisfies every target without narrowing the viewing scope. For video duration T_q, normalized excess access cost is

$$\text{NEC}_q = (|\widehat{I}_q \setminus E_q|)/(T_q - |E_q|). \tag{5}$$

NEC measures the fraction of the video's non-target content retained in the selected chapter. Normalizing by the available non-target duration means that a short target does not by itself imply high access cost. Successful Navigation Gain combines coverage success with reduced cost:

$$\text{SNG} = \frac{1}{|\mathcal{Q}_{\text{loc}}|} \sum_{q \in \mathcal{Q}_{\text{loc}}} H_q \left(1 - \text{NEC}_q\right), \tag{6}$$

where $\mathcal{Q}_{\text{loc}} = \{q : 0 < |E_q| < T_q\}$ contains legal queries with targets shorter than their videos. SNG rewards successful access that excludes non-target content and assigns zero to failures and whole-video selection. We report SNG alongside R@1; these duration-based measures approximate access cost, not observed viewing time or chapter coherence.

3.4 ChapterNavRL

Figure 3.1 (a) illustrates ChapterNav. A policy generates a reusable chapter directory, while a frozen navigator uses its titles to select an interval for each query. Verified navigation outcomes guide GRPO. We next describe the task, navigation supervision, evaluation, and training.

ChapterNavRL trains the generator using feedback from the navigation protocol. Starting from a supervised fine-tuned policy π_{θ_0}, we sample a group of G candidate directories from the current policy $\pi_\theta(\cdot \mid X)$ for each training video. A frozen navigator ranks their titles for the same fixed training queries, and target annotations verify selections to compute the reward in Section 3.5. Queries and target episodes are used exclusively for feedback computation and are never supplied to the chapter generator; evaluation queries are excluded from policy optimization.

Group Relative Policy Optimization (GRPO) DeepSeek-AI (2025) updates the generator using relative advantages computed from rewards within each candidate group. A clipped policy objective encourages higher-reward generations, while KL regularization against the frozen SFT policy π_{θ_0} constrains policy drift. The appendix specifies the objective, advantage normalization, sampling, and hyperparameters. Inference requires only the generator to produce a reusable directory.

3.5 Reward Design

The navigation reward combines first-choice gain with discounted feedback for later qualifying chapters. For local target E_q and video duration T, define $s_q = H_q(1 - NEC)$ as the query-level SNG contribution. Let $k_q^\star$ be the rank of the first chapter satisfying $\text{tR}_{qj} \geq \tau$; we define

$$d_q = \begin{cases} g_{q,\rho_q(k_q^\star)}/k_q^\star, & \text{if a qualifying chapter exists,} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Table 1: Reference agreement **(Ref.↑ = F1/tIoU/CIDEr against the ground truth)** and navigation utility **(Nav.↑ = R@1/SNG×100 at τ=0.9, MiniLM navigator; † the KZKT navigator is Qwen3-Embedding-4B, as MiniLM is unreliable on Chinese titles)**; all metrics are higher-is-better, and Nav. comparisons are made at full precision. L = Llama, Q = Qwen. **Bold**: R@1/SNG above the reference directory. Underlined: some system in the same column with strictly lower Ref. metrics achieves strictly higher R@1 *and* SNG.

	AVLecture		KZKT		SML300-val	
Model	Ref.↑	Nav.↑	Ref.↑	Nav†.↑	Ref.↑	Nav.↑
Zero-shot						
L3.2-3B	0.7/22.2/1.1	1.5/0.5	7.3/49.1/0.9	2.5/0.6	10.2/44.7/11.4	12.7/2.2
L3.1-8B	14.4/44.6/20.8	14.6/12.9	18.3/55.9/2.7	11.9/10.9	21.7/56.9/33.2	40.3/16.8
Q3-4B	13.2/48.8/27.5	25.3/23.8	14.7/54.6/9.8	17.8/16.6	23.5/57.4/33.1	41.6/22.7
Q3-8B	10.1/44.5/23.7	16.8/15.6	16.9/57.6/13.9	12.5/11.7	26.5/58.4/41.3	38.3/17.9
Q3.5-4B	39.3/71.3/92.1	53.4/46.5	39.6/71.6/43.2	47.1/41.0	36.6/67.3/50.7	53.9/23.2
Q3.5-9B	44.0/73.4/99.6	55.4/49.2	43.0/73.0/44.9	48.6/42.5	39.0/68.7/48.9	55.9/23.4
Chapter-LLaMa (SFT)						
L3.2-3B	47.6/60.2/107.8	60.9/46.7	60.1/79.7/106.8	48.6/35.6	34.7/65.1/65.0	30.3/24.9
L3.1-8B	49.6/63.5/118.2	60.1/48.8	63.0/81.1/124.8	**52.3**/36.4	41.2/69.7/79.4	38.7/33.1
Q3-4B	53.0/66.7/114.1	**64.9**/51.4	63.2/81.2/140.4	**51.9**/35.4	36.2/66.0/75.9	31.6/27.7
Q3-8B	44.7/73.0/127.3	52.5/46.4	64.9/82.3/145.2	**51.5**/34.8	32.6/62.4/74.4	31.0/27.2
Q3.5-4B	55.0/79.2/205.6	59.9/50.6	65.3/82.1/148.1	**52.9**/36.3	41.4/69.2/82.5	34.4/30.3
Q3.5-9B	49.0/76.4/150.4	**65.4**/**55.7**	66.3/82.9/149.6	**51.2**/38.1	44.1/70.8/119.3	37.1/33.1
API / Zero-shot						
`DS-V4-Fl`	24.4/59.4/54.7	54.1/23.9	18.6/58.8/81.4	31.2/16.1	38.6/69.2/51.7	**59.6**/26.6
`GPT-5.4 mini`	21.4/56.6/60.1	59.2/45.2	14.2/52.6/62.6	23.1/15.4	40.6/70.3/60.6	**61.2**/27.7
Reference GT	–/–/–	64.3/54.4	–/–/–	50.9/35.4	–/–/–	45.4/31.2

This auxiliary term provides feedback when first-choice navigation fails while retaining the access-efficiency criterion.

We average query rewards and gate the result by output validity. For fixed legal local training queries $\mathcal{Q}_v^{\text{loc}}$,

$$R_{\text{nav}}(C) = \frac{1}{|\mathcal{Q}_v^{\text{loc}}|} \sum_{q \in \mathcal{Q}_v^{\text{loc}}} (0.8 s_q + 0.2 d_q)\,. \tag{8}$$

First-choice success earns its access gain; otherwise, only discounted feedback contributes. Non-qualifying chapters earn no corresponding gain, and whole-video single-chapter directories receive zero navigation reward. Videos without legal local queries are excluded before training. The final reward is

$$R(C) = V_{\text{fmt}}(C) R_{\text{nav}}(C). \tag{9}$$

The binary gate checks the output schema, nonempty titles, and valid, strictly ordered timestamps forming a complete partition. Invalid outputs receive zero without component evaluation; valid outputs use the fixed weights above. Format checks and boundary cases appear in the appendix.

4 Experiments

4.1 Experimental Setup

Datasets. We conduct two groups of experiments: benchmark evaluation of existing chaptering systems and navigation-feedback learning for ChapterNavRL. The benchmark study uses three navigation-annotated evaluation sets: AVLecture-test, KZKT-test, and SML300-val. AVLecture-test contains 105 videos with 876 query-episode pairs, KZKT-test contains 436 videos with 3,868 pairs, and SML300-val contains 250 videos with 1,382 pairs. For navigation-feedback learning, ChapterNavRL is trained on the navigation-annotated training splits constructed in this work, including AVLecture-train with 245 videos and 2,049 query-episode pairs and KZKT-train with 757 videos

and 6,542 pairs. The detailed dataset construction and training split information are provided in the appendix.

Implementation Details. Our benchmark evaluates Llama-3.2-3B, Llama-3.1-8B, Qwen3-4B, and Qwen3.5-4B under zero-shot and SFT settings, together with API-based LLMs in the zero-shot setting. All models share the same input, prompt, output format, and parsing protocol. We use deterministic decoding with maximum input and output lengths of 16,384 and 2,048 tokens, respectively. Generated directories are independently evaluated by MiniLM and Qwen3-Embedding-4B with maximum encoding lengths of 256 and 512 tokens. For ChapterNavRL, we initialize Llama-3.2-3B and Qwen3-4B from their reference-supervised SFT checkpoints and optimize them using GRPO on training videos containing valid query-episode pairs. The training uses four generations per prompt, a temperature of 0.7, a learning rate of 2×10^{-6}, a KL coefficient of 0.03, and 150 optimization steps. We report the final checkpoint performance averaged over three random seeds, without validation-based checkpoint selection; additional LoRA configurations are provided in the appendix.

Metrics. Reference agreement is measured by boundary F1, matched temporal IoU (tIoU), and CIDEr. Navigation utility is measured at the query level by R@1 and SNG, where R@1 is first-choice success and SNG is the mean first-choice access gain rather than the composite training reward. F1, tIoU, and R@1 are displayed on a 0-100 percentage scale; SNG is multiplied by 100 for readability, and CIDEr retains its original scale.

4.2 Benchmark Results and Evaluation Agreement

Table 1 shows that reference agreement and navigation utility are not interchangeable: SFT raises all three reference metrics for every backbone, yet navigation does not follow. On SML300-val, SFT improves F1/tIoU but *lowers* R@1 for the strongest zero-shot backbones (Qwen3.5-4B: 53.9→34.4; Qwen3.5-9B: 55.9→37.1), placing them below even the reference directory (45.4), whereas their zero-shot counterparts and both API systems (59.6 and 61.2) navigate *better than the ground truth itself*. Reversals persist even between systems with similar agreement: on AVLecture, Qwen3-4B (SFT) has strictly lower F1/tIoU/CIDEr than Qwen3.5-4B (SFT) (53.0/66.7/114.1 vs. 55.0/79.2/205.6) yet higher R@1 and SNG (64.9/51.4 vs. 59.9/50.6).

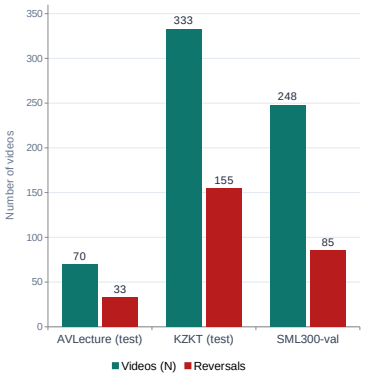

Figure 3: Per-video F1–SNG reversals between the zero-shot and SFT variants of Qwen3.5-4B.

We further quantify this dissociation at the system and video levels. At the system level, reference and navigation metrics are positively but imperfectly correlated (Pearson r from 0.19 for CIDEr vs. R@1 on SML300-val). At the video level, we count a *reversal* whenever SFT beats its zero-shot counterpart on F1 but not on SNG on the same video, or vice versa; for Qwen3.5-4B this occurs on 33/70 (AVLecture), 155/333 (KZKT), and 85/248 (SML300-val) videos (Figure 3). Reference metrics therefore capture coarse quality trends but are insufficient for evaluating navigation-oriented chapter generation.

Finally, we conduct a paired human evaluation on 50 videos from KZKT and AVLecture. Three annotators compare chapter plans generated by zero-shot and SFT Qwen3.5-4B and select the plan providing better navigation experience. The navigation-based metric agrees with human preference on 76% of pairs, compared with 68% for reference-based metrics, indicating that navigation utility provides a closer proxy to perceived chapter usability.

4.3 Comparison with Existing Chaptering Methods

We compare ChapterNavRL with two representative chaptering methods, Chapter-Llama (Ventura et al., 2025) and CausalChapter (Duan et al., 2026), on AVLecture and KZKT using Llama-3.2-3B and Qwen3-4B backbones. Both reference-based and navigation metrics are reported to evaluate chapter quality and downstream content access. ChapterNavRL is initialized from the corresponding

Table 2: Comparison with existing chaptering methods on AVLecture and KZKT. We report reference-based chaptering quality using F1, tIoU, and CIDEr, and navigation utility using R@1 and SNG. F1, tIoU, R@1, and SNG are reported in percentage points. For Avlecture we use MiniLM as Navigator while the Chinese dataset KZKT use Qwen3-4B-Embedding as the Navigator.

	AVLecture					KZKT				
	Reference-based			Navigation		Reference-based			Navigation	
Method	F1 ↑	tIoU ↑	CIDEr ↑	R@1 ↑	SNG ↑	F1 ↑	tIoU ↑	CIDEr ↑	R@1 ↑	SNG ↑
Chapter-Llama$^{\text{L3.2-3B}}$	47.6	60.2	107.8	60.9	46.7	60.1	79.7	106.8	48.6	35.6
Chapter-Llama$^{\text{Q3-4B}}$	53.0	66.7	114.1	64.9	51.4	63.2	81.2	140.4	51.9	35.4
CausalChapter$^{\text{L3.2-3B}}$	52.9	66.9	108.5	63.6	48.9	65.2	82.2	107.2	46.2	28.1
CausalChapter$^{\text{Q3-4B}}$	52.9	66.9	116.7	64.4	51.3	65.2	82.2	146.7	48.7	29.3
ChapterNavRL$^{\text{L3.2-3B}}$	49.1	61.5	111.3	65.8	51.4	61.3	80.6	108.8	50.3	36.9
ChapterNavRL$^{\text{Q3-4B}}$	55.1	69.2	144.2	68.6	54.8	65.7	82.9	142.8	53.1	36.8

Table 3: Effect of different rewards for GRPO post-training on AVLecture. All variants start from the same SFT checkpoint and differ only in the reward function. Ref-Temp and GRACE optimize reference-based temporal alignment, while Utility-RLVR optimizes downstream navigation utility.

		Reference-based			MiniLM Navigator			Qwen Navigator		
Backbone	Reward	F1 ↑	tIoU ↑	CIDEr ↑	R@1 ↑	SNG ↑	CondTR ↑	R@1 ↑	SNG ↑	CondTR ↑
L3.2-3B	SFT	47.6	60.2	107.8	60.9	46.7	61.9	64.1	49.4	65.2
	Ref-Temp	**49.5**	**62.1**	**114.3**	60.3	45.8	61.4	64.3	48.8	65.0
	GRACE	48.5	58.8	108.5	62.0	47.7	63.0	64.9	50.1	65.8
	Hybrid	49.3	61.9	112.2	61.2	46.9	62.2	63.7	48.8	64.5
	Utility-RLVR	49.1	61.5	111.3	**62.6**	**48.9**	**63.7**	**65.8**	**51.4**	**67.0**
Q3-4B	SFT	53.0	66.7	114.1	65.2	51.8	66.4	66.9	53.2	68.0
	Ref-Temp	54.4	68.1	**150.5**	64.1	50.5	65.1	65.6	52.4	66.6
	GRACE	54.9	68.8	139.9	65.9	52.3	67.0	67.4	54.0	68.5
	Hybrid	54.0	68.4	138.8	66.8	52.2	67.7	68.1	54.4	69.2
	Utility-RLVR	**55.1**	**69.2**	144.2	**67.6**	**53.7**	**68.5**	**68.6**	**54.8**	**69.6**

Chapter-Llama checkpoints, enabling a controlled comparison of navigation-guided reinforcement learning. CausalChapter is evaluated as a complete baseline because its segment-then-caption architecture optimizes chapter generation through a different modeling paradigm.

On AVLecture, ChapterNavRL consistently improves over its Chapter-Llama initialization under both backbones. With Llama-3.2-3B, R@1/SNG improves from 60.9/46.7 to 64.1/49.4, while Qwen3-4B improves from 65.2/51.8 to 68.6/54.8. The improvements in navigation metrics are accompanied by gains in tIoU and CIDEr, indicating that navigation-oriented post-training improves downstream utility without degrading conventional chapter quality. Compared with CausalChapter, ChapterNavRL$^{\text{Q3-4B}}$ achieves higher tIoU, CIDEr, R@1, and SNG, with only a slight decrease in F1, demonstrating that optimizing chapter generation with navigation feedback remains competitive with specialized chaptering architectures.

4.4 Learning from Verifiable Navigation Feedback

We investigate how reward design affects reinforcement learning for video chaptering on Avlecture dataset. Starting from the same SFT checkpoint, all GRPO variants use identical data, optimization configurations, and training budgets, differing only in reward functions. We compare our verifiable navigation reward (**Utility-RLVR**) with two reference-based temporal rewards: **Ref-Temp**, which optimizes temporal agreement with annotated chapters, and **GRACE** (Pu et al., 2025), which performs granularity-aware temporal matching. A **Hybrid** reward combining reference and navigation feedback is also evaluated. During Utility-RLVR training, MiniLM provides optimization feedback, while Qwen3-Embedding-4B is reserved as an independent evaluator.

Table 4: Sensitivity analysis of ChapterNavRL. The left table evaluates the verification threshold τ using the same generated directories without retraining. The right table studies the effect of the number of query–episode pairs sampled per video during RLVR training.

(a) Coverage Threshold τ

Metric	0.80	0.85	0.90	0.95	1.00
R@1 ↑	66.8	66.6	66.4	65.9	60.6
SNG ↑	53.2	53.1	52.9	52.4	47.5

(b) Queries per Video

Metric	1	2	4	8	ALL
R@1 ↑	65.3	65.9	66.1	66.6	66.4
SNG ↑	51.8	52.1	52.4	52.9	52.7

The results show that reward choice determines the behavior optimized by RL. Reference-based rewards improve agreement with annotations but do not necessarily improve navigation utility. For example, Ref-Temp achieves the highest CIDEr with Qwen3-4B, while its navigation performance decreases compared with the SFT initialization. In contrast, Utility-RLVR consistently improves navigation utility, increasing SNG from 51.8 to 53.7 under MiniLM evaluation and from 53.2 to 54.8 under the independent Qwen evaluator. GRACE provides moderate improvements by considering temporal granularity, but remains inferior to directly optimizing downstream utility. The Hybrid reward also fails to surpass Utility-RLVR, suggesting that reference-based supervision may introduce conflicting optimization signals. Furthermore, Utility-RLVR achieves the best F1 and tIoU among Qwen3-4B variants, showing that navigation-oriented optimization improves downstream usability without sacrificing conventional chapter quality.

4.5 Sensitivity Analysis

We analyze the sensitivity of ChapterNavRL to two factors that affect the navigation-based learning signal: the coverage threshold used for verification and the number of query–episode pairs sampled for each video during reinforcement learning.

Coverage Threshold. The main experiments use $\tau = 0.9$ to determine whether a selected chapter sufficiently covers the target episode. A larger τ imposes a stricter success criterion, while smaller values allow partial target coverage. We evaluate the same generated directories under different thresholds without retraining any model.

Number of Queries per Video. During RLVR training, multiple query-episode pairs provide navigation feedback for each video-level directory. We vary the number of sampled queries per video to study whether performance depends on reward density. All variants use the same optimization budget and training configuration.

As shown in Table 4, stricter coverage requirements reduce absolute navigation scores because fewer selected chapters satisfy the success criterion. However, performance trends remain stable around the default setting of $\tau = 0.9$, indicating that the effectiveness of ChapterNavRL is not sensitive to a specific verification threshold. Increasing the number of queries per video improves navigation performance by providing denser reward signals, while the improvement gradually saturates. This suggests that ChapterNavRL can obtain effective navigation feedback from a moderate number of queries without requiring excessive supervision.

5 Conclusion

Video chaptering is typically optimized toward agreement with reference organizations, but such agreement does not necessarily reflect whether a generated directory helps users access desired content. We introduce ChapterNav, a framework that evaluates and optimizes chapter generation through verifiable navigation utility, where downstream content access provides the learning signal for reinforcement learning with verifiable rewards. Experiments across multiple chaptering benchmarks demonstrate that navigation-oriented optimization improves practical access utility while maintaining competitive reference-based chapter quality, highlighting a new direction for evaluating and learning long-video structures beyond fixed reference annotations.

AI Use Statement

We used large language models (LLMs) in two aspects of this work. First, LLMs were used to polish the language of this manuscript (grammar and wording); all technical content, experimental design, and claims are the authors' own. Second, LLM agents assisted in constructing the navigation benchmarks introduced in Section **??**: following a protocol specified by human annotators, LLM agents inspected video content, discovered candidate episodes, drafted natural-language queries, and refined temporal boundaries.

References

Seung Hyup Baek, Jimin Lee, Hyeongkeun Lee, and Jae Won Cho. Stay in your lane: Role specific queries with overlap suppression loss for dense video captioning. *arXiv preprint arXiv:2603.11439*, 2026.

Satanjeev Banerjee and Alon Lavie. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pp. 65–72, 2005. URL https://aclanthology.org/W05-0909/.

DeepSeek-AI. DeepSeek-V3 technical report. *arXiv preprint arXiv:2412.19437*, 2024. URL https://arxiv.org/abs/2412.19437.

DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL https://arxiv.org/abs/2501.12948.

DeepSeek-AI et al. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025. URL https://arxiv.org/abs/2501.12948.

Xinran Duan, Guozhang Li, Yaoyao Zhong, Mei Wang, Lizhi Wang, and Hua Huang. Causalchapter: Improving long-video chaptering with interventional dependency modeling, 2026.

Soichiro Fujita, Tsutomu Hirao, Hidetaka Kamigaito, Manabu Okumura, and Masaaki Nagata. SODA: Story oriented dense video captioning evaluation framework. In *Computer Vision – ECCV 2020*, pp. 517–531, 2020. URL https://www.ecva.net/papers/eccv_2020/papers_ECCV/html/6406_ECCV_2020_paper.php.

Junyu Gao and Changsheng Xu. Fast video moment retrieval. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1503–1512. IEEE, 2021.

Anchit Gupta, CV Jawahar, Makarand Tapaswi, et al. Unsupervised audio-visual lecture segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5232–5241, 2023.

Minkuk Kim, Hyeon Bae Kim, Jinyoung Moon, Jinwoo Choi, and Seong Tae Kim. Do you remember? dense video captioning with cross-modal memory retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13894–13904, 2024.

Minkuk Kim, Hyeon Bae Kim, Jinyoung Moon, Jinwoo Choi, and Seong Tae Kim. Hicm^2: Hierarchical compact memory modeling for dense video captioning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 4293–4301, 2025.

Ranjay Krishna, Kenji Hata, Frederic Ren, Li Fei-Fei, and Juan Carlos Niebles. Dense-captioning events in videos. In *Proceedings of the IEEE international conference on computer vision*, pp. 706–715, 2017.

Nathan Lambert, Jacob Morrison, Valentina Pyatkin, Shengyi Huang, Hamish Ivison, Faeze Brahman, Lester James V. Miranda, Alisa Liu, Nouha Dziri, Shane Lyu, Yuling Gu, Saumya Malik, Victoria Graf, Jena D. Hwang, Jiangjiang Yang, Ronan Le Bras, Oyvind Tafjord, Chris Wilhelm, Luca Soldaini, Noah A. Smith, Yizhong Wang, Pradeep Dasigi, and Hannaneh Hajishirzi. Tülu 3: Pushing frontiers in open language model post-training. *arXiv preprint arXiv:2411.15124*, 2024. URL https://arxiv.org/abs/2411.15124.

Guozhang Li, Xinpeng Ding, De Cheng, Jie Li, Nannan Wang, and Xinbo Gao. Etc: Temporal boundary expand then clarify for weakly supervised video grounding with multimodal large language model. *IEEE Transactions on Multimedia*, 27:1772–1782, 2025.

Haoxuan Li, Mengyan Li, and Junjun Zheng. HiVid-Narrator: Hierarchical video narrative generation with scene-primed ASR-anchored compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Findings*, pp. 8195–8204, 2026. URL https://arxiv.org/abs/2601.07366.

Zihan Lin, Songhe Deng, Shuwei He, Danxiang Zhu, Dan Zhang, Yishu Lei, Xianlong Luo, Shikun Feng, and Rui Liu. CodecCap: High-fidelity codec-inspired residual modeling for dense video captioning. *arXiv preprint arXiv:2605.26967*, 2026. URL https://arxiv.org/abs/2605.26967.

WonJun Moon, Sangeek Hyun, SangUk Park, Dongchan Park, and Jae-Pil Heo. Query-dependent video representation for moment retrieval and highlight detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23023–23033. IEEE, 2023.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, 2002. doi: 10.3115/1073083.1073135. URL https://aclanthology.org/P02-1040/.

Junfu Pu, Teng Wang, Yixiao Ge, Yuying Ge, Chen Li, and Ying Shan. Arc-chapter: Structuring hour-long videos into navigable chapters and hierarchical summaries. *arXiv preprint arXiv:2511.14349*, 2025.

Yuxuan Qiao, Haodong Duan, Xinyu Fang, Junming Yang, Lin Chen, Songyang Zhang, Jiaqi Wang, Dahua Lin, and Kai Chen. Prism: A framework for decoupling and assessing the capabilities of VLMs. *arXiv preprint arXiv:2406.14544*, 2024. URL https://arxiv.org/abs/2406.14544.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. URL https://arxiv.org/abs/2402.03300.

Ramakrishna Vedantam, C. Lawrence Zitnick, and Devi Parikh. CIDEr: Consensus-based image description evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015. URL https://arxiv.org/abs/1411.5726.

Lucas Ventura, Antoine Yang, Cordelia Schmid, and Gül Varol. Chapter-llama: Efficient chaptering in hour-long videos with llms. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 18947–18958, 2025.

Teng Wang, Ruimao Zhang, Zhichao Lu, Feng Zheng, Ran Cheng, and Ping Luo. End-to-end dense video captioning with parallel decoding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6847–6857, 2021a.

Wenhui Wang, Hangbo Bao, Shaohan Huang, Li Dong, and Furu Wei. Minilmv2: Multi-head self-attention relation distillation for compressing pretrained transformers. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pp. 2140–2151, 2021b.

Long Xing, Xiaoyi Dong, Yuhang Zang, Yuhang Cao, Jianze Liang, Qidong Huang, Jiaqi Wang, Feng Wu, and Dahua Lin. CapRL: Stimulating dense image caption capabilities via reinforcement learning. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/15f1dbc086bfd94d8c32557b573cbe18-Abstract-Conference.html.

Antoine Yang, Arsha Nagrani, Ivan Laptev, Josef Sivic, and Cordelia Schmid. Vidchapters-7m: Video chapters at scale. *Advances in Neural Information Processing Systems*, 36:49428–49444, 2023a.

Antoine Yang, Arsha Nagrani, Paul Hongsuck Seo, Antoine Miech, Jordi Pont-Tuset, Ivan Laptev, Josef Sivic, and Cordelia Schmid. Vid2seq: Large-scale pretraining of a visual language model for dense video captioning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10714–10726, 2023b.

Penghui Yang, Long Xing, Xiaoyi Dong, Yuhang Zang, Yuhang Cao, Yibin Wang, Yujie Zhou, Jiazi Bu, Jianze Liang, Qidong Huang, Jiaqi Wang, Feng Wu, and Dahua Lin. CapRL++: Unified reinforcement learning with verifiable rewards for dense image and video captioning. *arXiv preprint arXiv:2606.09393*, 2026. URL https://arxiv.org/abs/2606.09393.

Shijia Yang, Yunong Liu, Bohan Zhai, Ximeng Sun, Zicheng Liu, Emad Barsoum, Manling Li, and Chenfeng Xu. CaptionQA: Is your caption as useful as the image itself? *arXiv preprint arXiv:2511.21025*, 2025. URL https://arxiv.org/abs/2511.21025.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*, 2025. URL https://arxiv.org/abs/2506.05176.

A Navigation Benchmark Construction

This appendix supplements Section 3.2 with the provenance, annotation procedure, and descriptive statistics of the two ChapterNav benchmarks. Both benchmarks contain model-proposed navigation queries paired with human-reviewed target episodes.

A.1 Data Sources and Splits

AVLecture. We use the AVLecture data version distributed through the MMVTS repository accompanying Yu et al. **?**.[1] The underlying AVLectures resource originates from the work of Singh et al. **?** on audio-visual lecture segmentation. The source topic-segmentation annotations specify topical boundaries; the navigation queries and their target episodes are additional annotations constructed for ChapterNav. Thus, the source segmentation annotations and our query-conditioned target episodes represent distinct annotation layers. Our retained AVLecture benchmark contains 350 videos, with 245/35/70 videos in the training/development/test splits.

KZKT. KZKT is drawn from teaching videos obtained from China's National Smart Education Platform for Primary and Secondary Schools. It covers multiple subjects within K-12 education. The retained benchmark contains 1,193 videos, with 757/99/337 videos in the training/development/test splits. We report the split membership stored in the final navigation release; these counts refer to the retained benchmark, not the larger collection from which videos were selected. All queries associated with a video belong to that video's split. We verify that video identifiers are disjoint across splits within eachbenchmark.

A.2 Reference Chapter Annotation

We constructchapter-level references for AVLecture and KZKT. Each reference chapter consists of a temporal interval, a concise title, and a content summary. These references support supervised training and reference-based evaluation of video chaptering models.

Temporal chapter annotation. For AVLecture, we retain the original topic-segmentation boundaries and supplement the resulting segments with chapter titles and summaries. For KZKT, the author team invited education professionals from a Normal university to manually delineate chapter segments based on the instructional content.

Title and summary annotation. We adapt the chapter-description annotation approach of CausalChapter (Duan et al., 2026) to the title–summary format used in this work. Given a chapter interval and its corresponding video content, the annotation model generates a concise title identifying the main topic and a summary describing the content within that interval. For both AVLecture and KZKT, education professionals from a Normal university subsequently review and revise the generated titles and summaries. This human proofreading and revision step produces the final chapter-level textual references.

Relation to navigation annotations. The reference chapters and navigation annotations serve different purposes. Reference chapters provide a video-level organization for chaptering supervision and reference-based evaluation, whereas query–episode pairs specify the content a viewer seeks to locate. A query-specific target episode therefore need not coincide with a reference chapter.

A.2.1 Navigation Queries and Target Episodes

A navigation query q expresses the content a viewer wishes to locate within a video, such as an explanation, derivation, worked example, or comparison. Its target episode is a contiguous interval $E_q = [a_q, b_q]$ in a video of duration T, with $0 \leq a_q < b_q \leq T$. The annotation objective is a compact, semantically complete interval containing sufficient content to satisfy the query. Compactness does not imply a mathematically shortest interval: necessary premises, intermediate reasoning, and a closing explanation may all belong to the target. Queries and episodes are linked one-to-one by

[1] `https://github.com/alibaba-damo-academy/SpokenNLP/tree/main/mmvts`

identifiers in the final release. Different queries may refer to overlapping content; the episodes are query-specific targets, not a non-overlapping partition ofthe video. In particular, a target episode need not coincide with a source topic segment. Queries are intended for subsequent navigation supervision and evaluation, rather than inputs to the query-independent chapter generator.

A.3 Dual-Model Construction and Cross-Verification

For both datasets, we use API-based annotation with GPT-5.5 and DeepSeek-v4 Flash to propose and cross-check navigation query-episode pairs. The construction associates a content-location request with supporting video content and candidate interval boundaries. Cross-verification checks whether the query and target describe the same content need and whether the proposed episode provides a coherent response to that need. Only candidates accepted by both models proceed to human review. Here, "intersection" refers to agreement on candidate acceptance, not to taking the temporal intersection of two proposed intervals.

Prompt documentation. The templates below are reconstructed reproduction templates that explicitly express the intended annotation criteria. Both templates can be applied to the corresponding dataset language.

> **Reconstructed candidate prompt.** You are annotating navigation needs for an educational video. Given its content and the available temporal alignment, propose a natural request that a viewer could use to locate a specific explanation, derivation, example, or comparison. Write the query in the video's language. Select one compact, contiguous episode that contains enough context and a complete explanation to satisfy that request. Do not require a chapter number, timestamp, or copied chapter title to make the query meaningful. Avoid whole-video summaries and requests requiring disconnected spans. Return the query, the proposed start and end, the supporting content references, and a short justification of sufficiency and boundary placement. Use only evidence in the supplied content; reject a candidate when the evidence is insufficient. Do not interpret estimated alignment as a manually verified video clock.

A.4 Human Review and Finalization

Approximately ten volunteers participated in human review. Every retained query-episode pair was reviewed by at least two people. Reviewers checked the correspondence between the query and its supporting episode, and samples judged unsuitable during human review were discarded. The counts reported here describe the final retained records after this process.

Reconstructed review checklist. It is a proposed formalization of the review process, not evidence that these exact questions were recorded for every reviewer.

1. **Naturalness and clarity:** Does the query describe a clear, specific content-location need in natural language?
2. **Correspondence and sufficiency:** Does the episode contain the information needed to satisfy that query?
3. **Semantic completeness:** Are necessary premises, explanation steps, and conclusions included?
4. **Compactness and continuity:** Is clearly irrelevant material avoided, and can the need be met in one continuous episode?
5. **Navigation suitability:** Is the request free of substantive ambiguity and dependence on timestamps or chapter numbers?

A.4.1 Benchmark Statistics

Table 5 reports the final video-level split sizes. A query-episode pair is one query and its linked episode, counted once; total video hours count each stored video duration once, regardless of the number or overlap of its episodes. Table 6 summarizes annotation granularity. Video duration and query count statistics weight each video equally, whereas episode duration and episode/video ratios

Table 5: Final ChapterNav benchmark sizes by video-level split. A pair is one navigation query and its linked target episode. Hours sum each video's stored duration once, not episode durations. Values are rounded independently.

Dataset	Split	Videos	Query–episode pairs	Hours
AVLecture	Train	245	2,049	207.5
AVLecture	Dev	35	280	29.9
AVLecture	Test	70	596	59.6
AVLecture	Total	350	2,925	297.0
KZKT	Train	757	6,542	350.8
KZKT	Dev	99	923	46.9
KZKT	Test	337	2,945	159.7
KZKT	Total	1,193	10,410	557.4

Table 6: Benchmark granularity over all splits. Video duration and query count are video-weighted; episode duration and its percentage of video duration are query-weighted. KZKT episode times use text-length-based estimated alignment, without per-episode video-clock verification.

	AVLecture		KZKT	
Measure	Mean	Median	Mean	Median
Video duration (min)	50.91	49.26	28.03	24.37
Queries per video	8.36	7.00	8.73	9.00
Episode duration (s)	206.77	175.00	141.36	129.82
Episode/video duration (%)	7.04	5.66	8.44	7.76

weight each retained query equally. Means and medians are computed over all three splits. Figure 4 complements these summaries with empirical distributions. Each ECDF uses all observations, including the upper tail; the query-count distribution is normalized separately by the number of videos in each dataset. Episode/video ratios are computed per pair before aggregation, rather than as the ratio of aggregate episode and video durations. These are descriptive statistics of the retained benchmark, not estimates of the prevalence of navigation needs among real users. The accompanying statistical validation checks nonempty queries, bidirectional query–episode references, unique video identifiers, split membership, and finite intervals satisfying $0 \leq a_q < b_q \leq T$. Both datasets pass these structural checks.

B Additional Details of Navigation Evaluation

This section provides additional implementation details and analyses for the navigation evaluation protocol introduced in Sec. **??**. We first describe the frozen title-based navigator, and then examine the robustness of the evaluation to navigator choice and the target-coverage threshold. We further decompose navigation performance into structural availability, title-based routing, and access efficiency, and discuss the properties and boundary cases of Successful Navigation Gain (SNG).

B.1 Navigator Implementation

The navigation protocol intentionally restricts the navigator to information exposed by the generated directory. Given a navigation query q and a generated chapter title t_j, a frozen text encoder $\phi(\cdot)$ produces their representations

$$\mathbf{z}_q = \phi(q), \qquad \mathbf{z}_j = \phi(t_j). \tag{10}$$

We compute the query–title relevance score using cosine similarity,

$$r_{qj} = \frac{\mathbf{z}_q^\top \mathbf{z}_j}{||\mathbf{z}_q||_2 ||\mathbf{z}_j||_2}, \tag{11}$$

and rank all chapters in descending order of $r * qj$:

$$\rho_q = \text{argsort}_j, r_{qj}. \tag{12}$$

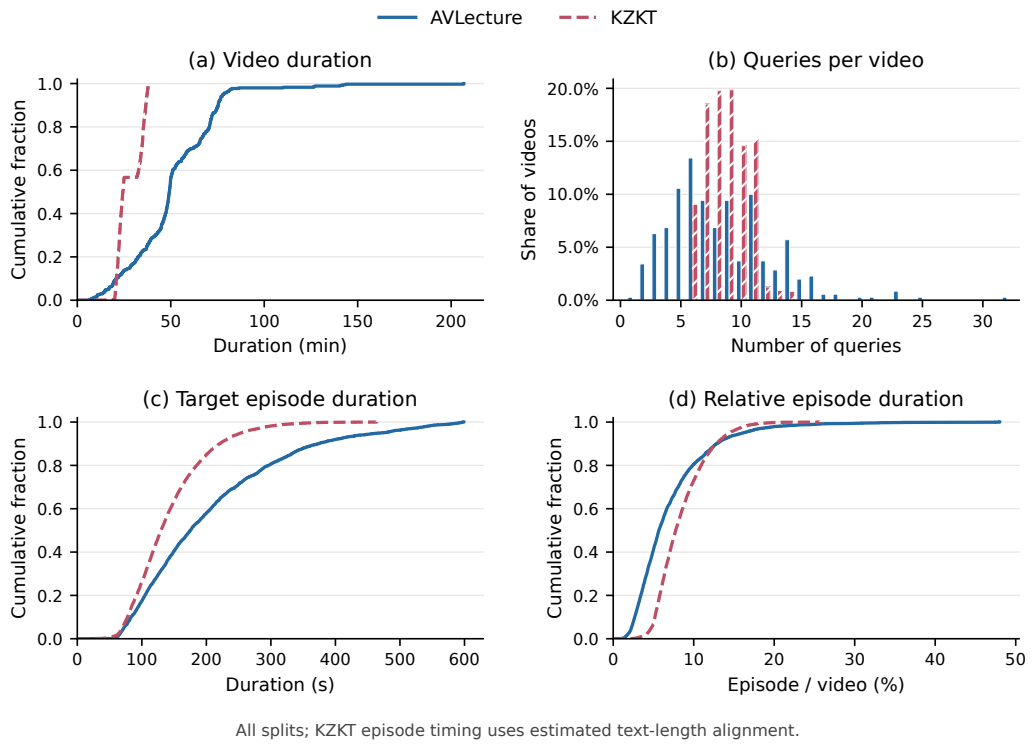

Figure 4: Distributions in the final ChapterNav benchmarks, pooling training, development, and test splits. (a) Video duration, with each video weighted equally. (b) Number of queries per video, normalized within each dataset. (c) Target episode duration. (d) Episode duration as a percentage of its video's duration. Panels (c) and (d) weight query-episode pairs equally. AVLecture contains 350 videos and 2,925 pairs; KZKT contains 1,193 videos and 10,410 pairs. KZKT episode times use text-length-based estimated alignment and were not individually verified against the video playback clock.

The top-ranked interval is therefore $\widehat{I}_q = I_{\rho_q(1)}$.

Unless otherwise specified, we use `Qwen3-4B-Embedding` as the default navigator. The navigator remains frozen throughout both training and evaluation and receives only the query and generated chapter titles. It has no access to the video representation, transcript, target episode, reference directory, or hidden states of the chaptering model. The same navigator and ranking configuration are used for all compared chaptering systems.

Embeddings are ℓ_2-normalized before similarity computation. Queries and chapter titles are encoded independently, and no generator-specific adaptation is applied to the navigator. In the rare case of identical relevance scores, ties are resolved according to the temporal order of the chapters. These restrictions ensure that differences in navigation performance arise from information exposed by the generated directories rather than additional access to the underlying video.

B.2 Decomposing Navigation Performance

The navigation protocol separates two distinct requirements for successful content access: the generated temporal partition must first contain an appropriate chapter, and the generated title must then allow the navigator to select it. The indicators

$$A_q = \mathbb{1}\left[\max_j \mathrm{tR} * qj \geq \tau\right], \qquad H_q = \mathbb{1}\left[\mathrm{tR} * q, \rho_q(1) \geq \tau\right] \tag{13}$$

make this distinction explicit.

Table 7: Decomposition of navigation performance on Avlecture. BestTR measures whether an appropriate chapter exists, CondTR measures successful title routing conditional on availability, and NEC_{succ} measures excess access cost among successful selections. SNG jointly captures navigation success and access efficiency.

Method	BestTR ↑	CondTR ↑	R@1 ↑	NEC_{succ} ↓	SNG ↑
SFT	98.5	66.9	68.0	XX	53.2
Reference-RL	98.5	65.6	66.6	20.0	52.2
ChapterNavRL	98.7	69.6	68.5	20.2	54.8

BestTR measures *structural availability*: whether at least one generated chapter sufficiently covers the target episode. CondTR measures *title routing*: whether the navigator selects such a chapter when one is available. R@1 measures their end-to-end combination. From the definitions in Sec. 3.3,

$$\text{BestTR} = \frac{\sum_q A_q}{N}, \quad \text{CondTR} = \frac{\sum_q H_q}{\sum_q A_q}, \tag{14}$$

and therefore, whenever $\sum_q A_q > 0$,

$$\text{R@1} = \text{BestTR} \times \text{CondTR}. \tag{15}$$

This identity allows end-to-end navigation failures to be attributed to either the absence of an appropriate temporal unit or unsuccessful title-based routing.

We additionally report excess access cost only over successful navigation cases:

$$\text{NEC}_{\text{succ}} = \frac{\sum_q H_q \, \text{NEC}_q}{\sum_q H_q}. \tag{16}$$

Conditioning on $H_q = 1$ avoids assigning apparently favorable access cost to short but incorrect chapter selections.

The decomposition provides complementary information to aggregate SNG. Improvements in BestTR indicate that the generated temporal partition more frequently exposes a chapter containing the required content. Improvements in CondTR indicate that chapter titles more effectively identify these intervals, while reductions in NEC_{succ} indicate that successful navigation requires viewing less unrelated content. Together, these quantities reveal whether gains in SNG arise from improved temporal organization, better chapter naming, more efficient access, or a combination of these factors.

C Properties and Budget Interpretation of SNG

We next discuss several properties of SNG that motivate its use as the primary navigation metric. For every legal local query,

$$0 < |E_q| < T_q. \tag{17}$$

Since the selected chapter is contained within the video,

$$0 \le |\widehat{I}_q \setminus E_q| \le T_q - |E_q|, \tag{18}$$

which implies

$$0 \le \text{NEC}_q = \frac{|\widehat{I}_q \setminus E_q|}{T_q - |E_q|} \le 1. \tag{19}$$

The query-level SNG contribution

$$s_q = H_q \, (1 - \text{NEC}_q) \tag{20}$$

is therefore also bounded:

$$0 \le s_q \le 1, \qquad 0 \le \text{SNG} \le 1. \tag{21}$$

Ideal access. If the selected chapter exactly matches the target episode,

$$\widehat{I}_q = E_q, \tag{22}$$

then $\mathrm{NEC}_q = 0$. Since the target is completely covered, $H_q = 1$, yielding

$$s_q = 1. \tag{23}$$

Thus, the maximum score corresponds to successful access with no unnecessary temporal content.

Whole-video degeneracy. Target recall alone admits a trivial solution in which the complete video is exposed as a single chapter. If

$$\widehat{I}_q = [0, T_q], \tag{24}$$

then target coverage is complete, but

$$\mathrm{NEC}_q = \frac{T_q - |E_q|}{T_q - |E_q|} = 1, \tag{25}$$

and hence

$$s_q = 0. \tag{26}$$

SNG therefore does not reward a directory that obtains perfect target recall simply by requiring the user to inspect the entire video.

Failed navigation. If the top-ranked chapter does not satisfy the target-coverage requirement, then $H_q = 0$ and

$$s_q = 0 \tag{27}$$

regardless of chapter length. A short but incorrect chapter therefore cannot obtain a favorable navigation score merely because it has low access cost.

Monotonic access efficiency. For successful selections with fixed target episode E_q, decreasing the amount of non-target content monotonically increases the query-level score:

$$|\widehat{I}_q \setminus E_q| \downarrow \implies \mathrm{NEC}_q \downarrow \implies s_q \uparrow . \tag{28}$$

SNG therefore favors tighter access intervals only after the target-coverage requirement has been satisfied.

Viewing-budget interpretation. SNG can also be interpreted as navigation success aggregated over normalized viewing budgets. Let

$$b \in [0, 1] \tag{29}$$

denote the maximum normalized excess-access cost a user is willing to tolerate, and define success under budget b as

$$S_q(b) = H_q \mathbb{1}\left[\mathrm{NEC}_q \le b\right]. \tag{30}$$

A query succeeds under budget b only when the navigator first selects a qualifying chapter and the excess content contained in that chapter remains within the allowed budget.

Integrating this success indicator over all normalized budgets gives

$$\begin{aligned} \int_0^1 S_q(b)\, db &= H_q \int_0^1 \mathbb{1}\left[\mathrm{NEC}_q \le b\right] db \\ &= H_q \left(1 - \mathrm{NEC}_q\right) \\ &= s_q. \end{aligned} \tag{31}$$

Hence, the query-level SNG contribution is equivalently the area under the navigation-success curve over normalized excess-viewing budgets. Averaging over all legal local queries yields

$$\mathrm{SNG} = \frac{1}{|\mathcal{Q}_{\mathrm{loc}}|} \sum_{q \in \mathcal{Q}_{\mathrm{loc}}} \int_0^1 S_q(b)\, db. \tag{32}$$

This interpretation connects target coverage and access efficiency without selecting a single arbitrary viewing-budget threshold: a directory obtains a high SNG only when it supports successful navigation across a broad range of increasingly restrictive access budgets.

D ChapterNavRL Training Details

This section provides additional training details for ChapterNavRL. We use the supervised video chaptering model following Chapter-Llama Ventura et al. (2025) as the initialization and subsequently optimize it with Group Relative Policy Optimization (GRPO) using the navigation reward defined in Sec. 3.5.

D.1 Supervised Initialization

We first obtain a supervised policy π_{θ_0} following the training formulation of Chapter-Llama Ventura et al. (2025). Given the time-aligned textual video representation X, the model is trained to autoregressively generate the reference chapter directory, including chapter boundaries and titles. The resulting SFT checkpoint serves as both the initialization of the trainable policy and the reference policy used to regularize reinforcement learning.

Unless otherwise specified, all ChapterNavRL variants start from the same SFT checkpoint. We therefore attribute differences after reinforcement learning to the optimization objective rather than differences in supervised initialization.

D.2 GRPO Optimization

Starting from π_{θ_0}, we sample a group of G candidate directories

$$C_1, \ldots, C_G \sim \pi_\theta(\cdot \mid X) \tag{33}$$

for each training video. Each candidate directory is independently evaluated using the frozen navigator and the fixed training queries associated with the video, producing rewards

$$R_1, \ldots, R_G. \tag{34}$$

Queries and target episodes are used only for reward computation and are never included in the input to the chapter generator.

Following GRPO, we normalize rewards within each candidate group to obtain relative advantages:

$$A_i = \frac{R_i - \mu_R}{\sigma_R + \epsilon}, \tag{35}$$

where

$$\mu_R = \frac{1}{G}\sum_{i=1}^{G} R_i, \qquad \sigma_R = \sqrt{\frac{1}{G}\sum_{i=1}^{G}(R_i - \mu_R)^2}. \tag{36}$$

Here, ϵ is a small constant for numerical stability. When all candidates in a group receive the same reward, their relative advantages are zero and the group does not contribute a preference signal.

The policy is updated with the standard clipped GRPO objective while regularizing deviations from the frozen SFT policy π_{θ_0}:

$$\mathcal{L}_{\text{GRPO}} - \frac{1}{G}\sum_{i=1}^{G}\Big[\min\left(r_i(\theta)A_i, , \text{clip}(r_i(\theta), 1-\epsilon_{\text{clip}}, 1+\epsilon_{\text{clip}})A_i\right) \quad -\beta D_{\text{KL}}\left(\pi_\theta \,\|\, \pi_{\theta_0}\right)\Big], \tag{37}$$

where $r_i(\theta)$ denotes the policy ratio for candidate C_i, ϵ_{clip} controls policy clipping, and β controls KL regularization. We follow the same GRPO implementation for all reinforcement-learning baselines and vary only the reward being optimized.

D.3 Training Configuration

Table 8 summarizes the main reinforcement-learning configuration. Candidate directories are sampled stochastically during training, whereas inference uses the same decoding configuration as the corresponding SFT model unless otherwise stated. No navigator or verifier is required at inference time.

Table 8: Main training hyperparameters for ChapterNavRL.

Hyperparameter	Value
SFT initialization	Chapter-Llama checkpoint
Group size G	4
Learning rate	2×10^{-6}
Training epochs / steps	300 steps ($\sim$8.8 / 1.7 epochs on AVL / KZKT)
Global batch size	16 completions (4 prompts $\times$ G)
Gradient accumulation	16
Sampling temperature	0.7
Top-p	1.0
Maximum output length	512 tokens
GRPO clip range ϵ_{clip}	0.2
KL coefficient β	0.03
Optimizer	AdamW
Precision	BF16 (4-bit base, LoRA r=16)

E Additional Analysis of Reward Design

This section provides additional analysis of the navigation reward introduced in Sec. 3.5. We first clarify its relation to the evaluation objective, and then study the role of rank-based reward shaping and the sensitivity to its weighting. The reward is designed so that successful top-ranked navigation receives exactly the same query-level gain used by SNG, while additional shaping is introduced only when a useful chapter exists but is not ranked first.

E.1 Relation between Navigation Reward and SNG

For a query q and chapter I_j, recall the access gain

$$g_{qj} = 1 - \frac{|I_j \setminus E_q|}{T - |E_q|}. \tag{38}$$

For the top-ranked chapter, the primary navigation reward is

$$s_q = H_q g_{q,\rho_q(1)}. \tag{39}$$

This is exactly the query-level contribution used to compute SNG:

$$s_q = H_q \left(1 - \text{NEC}_q\right). \tag{40}$$

Consequently, the primary training objective directly optimizes the same notion of successful and efficient content access used for evaluation rather than a separate surrogate based on reference similarity.

The auxiliary rank-based term is defined using the first qualifying chapter in the navigator ranking. Specifically, we define

$$k_q^\star = \min\left\{k \,\middle|\, \text{tR}_{q,\rho_q(k)} \geq \tau\right\}, \tag{41}$$

and set $k_q^\star = \infty$ if no qualifying chapter exists. The auxiliary reward is then defined as

$$d_q = \begin{cases} \dfrac{g_{q,\rho_q(k_q^\star)}}{k_q^\star}, & k_q^\star < \infty, \\ 0, & k_q^\star = \infty. \end{cases} \tag{42}$$

The per-query training reward is defined as

$$r_q(\alpha) = \alpha s_q + (1-\alpha) d_q, \tag{43}$$

where $\alpha = 0.8$ in our main experiments.

An important property is that the auxiliary term does not alter the reward when top-ranked navigation already succeeds. If the top-ranked chapter qualifies, then

$$k_q^\star = 1, \tag{44}$$

Table 9: Ablation of navigation reward components on Avlecture. Coverage-only reward removes the access-efficiency term, while SNG-only reward removes rank-based shaping. All variants start from the same SFT checkpoint and use the same GRPO configuration.

Reward	BestTR ↑	CondTR ↑	R@1 ↑	NEC_{succ} ↓	SNG ↑
Coverage only R_{cov}	98.8	66.7	65.9	21.9	52.2
SNG only R_{SNG}	98.3	67.6	66.4	21.4	52.8
Full reward R_{full}	98.8	68.9	68.1	20.3	54.8

and consequently

$$s_q = d_q = g_{q,\rho_q(1)}. \tag{45}$$

Therefore,

$$r_q(\alpha) = g_{q,\rho_q(1)} = s_q, \qquad \forall\, \alpha \in [0,1]. \tag{46}$$

Equivalently, the per-query reward can be written as

$$r_q(\alpha) = \begin{cases} g_{q,\rho_q(1)}, & k_q^\star = 1, \\ (1-\alpha)\dfrac{g_{q,\rho_q(k_q^\star)}}{k_q^\star}, & 1 < k_q^\star < \infty, \\ 0, & k_q^\star = \infty. \end{cases} \tag{47}$$

Thus, rank-based shaping leaves the original SNG reward unchanged for successful top-ranked navigation, while providing a weaker auxiliary signal when a qualifying chapter already exists but is ranked below the first position.

E.2 Effect of Rank-Based Reward Shaping

Using only s_q gives zero reward whenever the navigator fails to select a qualifying chapter at rank one. This includes two qualitatively different situations: the generated directory may contain no chapter that sufficiently covers the target, or it may already contain an appropriate chapter whose title is ranked below another entry. Treating both situations identically makes the reward unnecessarily sparse.

The auxiliary term d_q distinguishes these cases. A directory with a qualifying chapter at rank $k_q^\star > 1$ receives a small positive reward proportional to its access gain and inversely proportional to its rank. Directories containing no qualifying chapter still receive zero. The rank discount ensures that this auxiliary feedback remains substantially weaker than successful top-ranked navigation.

We evaluate whether this shaping term improves optimization by comparing the following reward variants:

$$R_{\text{cov}} = \frac{1}{|\mathcal{Q}_v^{\text{loc}}|}\sum_q H_q, \qquad R_{\text{SNG}} = \frac{1}{|\mathcal{Q}_v^{\text{loc}}|}\sum_q s_q, \quad R_{\text{full}} = \frac{1}{|\mathcal{Q}_v^{\text{loc}}|}\sum_q \left(0.8 s_q + 0.2 d_q\right). \tag{48}$$

R_{cov} verifies successful target coverage but does not penalize unnecessary viewing, allowing us to isolate the contribution of access efficiency. R_{SNG} directly optimizes the evaluation objective without auxiliary shaping, while R_{full} is the reward used by ChapterNavRL.

The comparison separates two effects of the proposed reward. Comparing R_{cov} with R_{SNG} evaluates whether explicitly penalizing excess access encourages more efficient chapter structures. Comparing R_{SNG} with R_{full} evaluates whether rank-based shaping improves optimization beyond the primary navigation objective.

E.3 Sensitivity to the Reward Weight

The main experiments use $\alpha = 0.8$:

$$r_q = 0.8 s_q + 0.2 d_q. \tag{49}$$

Table 10: Sensitivity to the primary-reward weight α in $r_q(\alpha) = \alpha s_q + (1-\alpha)d_q$. All other training configurations are unchanged.

α	R@1 ↑	NEC_{succ} ↓	SNG ↑	Informative groups ↑
1.0	67.11	20.12	53.61	91.92
0.9	68.62	20.09	54.84	92.58
0.8	67.62	20.12	54.01	92.50

Table 11: Fraction of invalid directory outputs at different stages of training.

Model / Training Stage	Invalid outputs (%) ↓
SFT initialization	2.94
ChapterNavRL, early training	10.12
ChapterNavRL, final checkpoint	8.25

Because $s_q = d_q$ whenever top-ranked navigation succeeds, α controls only the magnitude of auxiliary feedback for unsuccessful top-ranked navigation. We therefore expect the method to be relatively insensitive to moderate changes in this coefficient.

We evaluate

$$\alpha \in \{1.0, 0.9, 0.8\}, \tag{50}$$

where $\alpha = 1$ corresponds to removing rank-based shaping entirely.

The results show whether improvements require precise tuning of the shaping coefficient. The default $\alpha = 0.8$ is selected to keep successful top-ranked navigation as the dominant training objective while providing a small amount of feedback for useful chapters appearing lower in the ranking.

E.4 Output Validity Gate

The final reward additionally applies the binary validity gate

$$R(C) = V_{\text{fmt}}(C) R_{\text{nav}}(C). \tag{51}$$

The gate verifies that the generated output can be interpreted as a valid chapter directory. Specifically, it checks that the output follows the required schema, contains non-empty chapter titles, uses valid and strictly ordered timestamps, and forms a complete partition of the video timeline.

The validity gate is not intended to express chapter quality. Instead, it prevents malformed outputs from obtaining navigation rewards based on partially parsed or ambiguous structures. Invalid outputs receive zero reward and are not repaired before reward computation.

Rather than treating this deterministic validity requirement as a competing reward objective, we report the fraction of invalid sampled directories during training.

This analysis verifies that navigation improvements are obtained primarily through changes in valid chapter structures rather than through the format gate itself.

F Human Navigation Preference Study

We conduct a small-scale human study to examine whether the proposed navigation metric better reflects the practical utility of generated chapter directories. Rather than asking annotators to judge agreement with a reference chapter structure, we directly ask which of two generated directories would be more useful for locating content in the corresponding video.

F.1 Human Preference Protocol

We sample 50 videos and construct one pair of candidate chapter directories for each video from outputs produced by different chaptering models. For every comparison, annotators are shown the

Table 12: Agreement between automatic evaluation and human preferences over generated chapter directories. Higher agreement indicates that the metric more closely reflects which directory humans consider more useful for content access.

Evaluation criterion	Agreement with human preference (%) ↑
Reference-based evaluation	68.0
SNG	76.0

video together with two generated chapter directories, denoted as $C^{(A)}$ and $C^{(B)}$. Each directory contains only the generated chapter boundaries and titles shown to an end user.

Annotators are asked to compare the two directories from the perspective of practical content access and select the one that would better help them locate desired information in the video. They are instructed to consider whether the chapter titles provide informative entry points, whether the temporal organization makes relevant content easy to locate, and whether the resulting chapters avoid requiring excessive viewing of unrelated content.

The identities of the generating systems are hidden from annotators, and the presentation order of the two directories is randomized. Reference chapter annotations and automatic metric scores are not shown. Each pair is independently judged by multiple annotators, and the final human preference is determined by majority vote. Tied or unresolved comparisons are excluded from the agreement calculation.

In total, the study contains 50 directory pairs evaluated by 5 annotators. Annotator agreement is XX, measured by Fleiss' κ.

F.2 Agreement with Automatic Metrics

We compare human pairwise preferences with the preferences induced by automatic evaluation metrics. For a metric m, let $m(C^{(A)})$ and $m(C^{(B)})$ denote the scores assigned to the two candidate directories. The metric-induced preference is defined as

$$P_m = \begin{cases} A, & m(C^{(A)}) > m(C^{(B)}), \\ B, & m(C^{(B)}) > m(C^{(A)}), \\ \text{tie}, & m(C^{(A)}) = m(C^{(B)}). \end{cases} \tag{52}$$

Let P_h denote the corresponding majority human preference. We measure metric-human agreement as

$$\text{Agree}(m) = \frac{\sum_{i=1}^{N} \mathbb{1}\left[P_m^{(i)} = P_h^{(i)}\right]}{N}, \tag{53}$$

where N is the number of valid pairwise comparisons after excluding unresolved human judgments and metric ties.

We apply the same protocol to SNG and reference-based chapter evaluation, so that both metrics are compared against identical human judgments.

SNG shows higher agreement with human navigation preferences than reference-based evaluation. This result supports our central motivation that similarity to a particular reference organization is not always aligned with the practical content-access utility of a chapter directory, while navigation-based evaluation provides a closer approximation to human judgments of usefulness.

G Additional Qualitative Analysis

We provide additional qualitative examples to illustrate the difference between reference agreement and navigation utility, the behavioral changes induced by ChapterNavRL, and representative failure cases. These examples complement the aggregate results by showing how different chapter organizations affect practical content access.

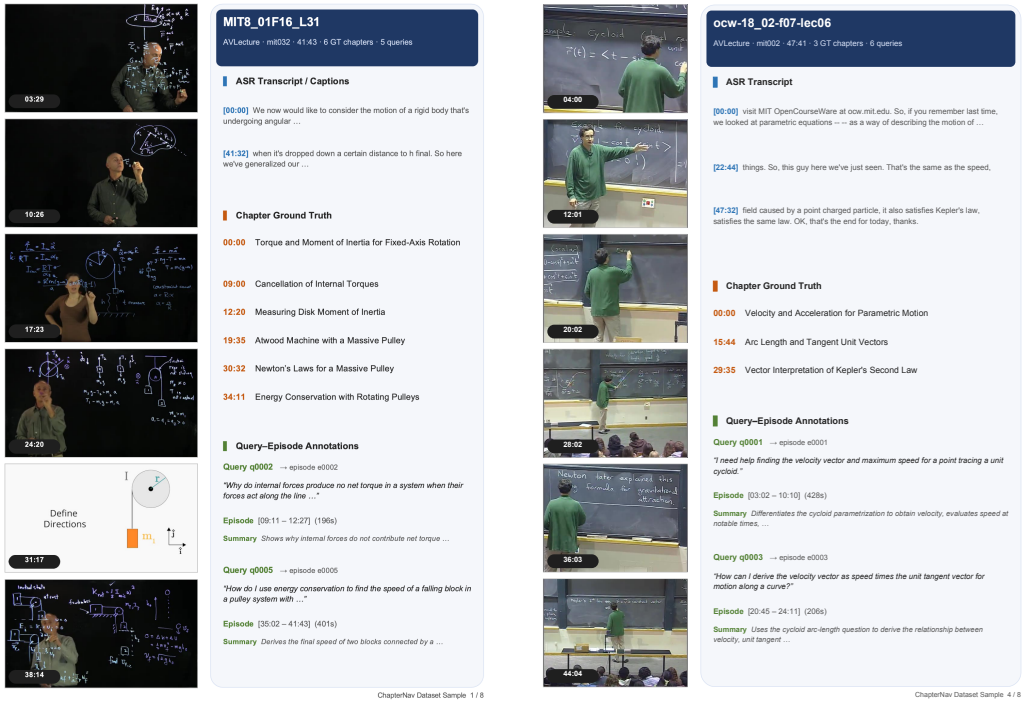

Figure 5: Representative ChapterNav samples from a physics lecture (left) and a mathematics lecture (right). Each showcase presents video metadata, timestamped frames, transcript excerpts, reference chapters, and selected query–episode annotations. Each displayed query is paired with a target temporal interval and a brief content summary, illustrating the annotation structure used for downstream navigation utility evaluation.

G.1 ChapterNav Dataset Examples

Figure 5 presents two representative samples from ChapterNav, our dataset for evaluating the downstream navigation utility of video chapter directories. Each showcase includes timestamped video frames, excerpts from the ASR transcript or captions, reference chapter titles with their start times, and selected query–episode annotations. The sample header reports the video identifier, source, duration, and numbers of reference chapters and navigation queries. Together, these components illustrate how a video's chapter structure is associated with specific information needs and their corresponding temporal targets.

Each navigation query expresses an information need in natural language and is linked to a target episode annotated with its start and end times, duration, and a brief content summary. In the physics lecture, for example, a query about the cancellation of internal torques is associated with the interval from 09:11 to 12:27. In the mathematics lecture, a query about the velocity vector and maximum speed of a point tracing a unit cycloid is linked to the interval from 03:02 to 10:10. These examples illustrate the complementary roles of the annotations: reference chapters describe the video's topical organization, while query–episode pairs specify the content that a navigator should reach. ChapterNav thereby supports evaluating how effectively a generated chapter directory helps users locate content relevant to a concrete information need.

G.2 SFT versus ChapterNavRL

We further compare directories generated by the common SFT initialization and ChapterNavRL. Figure 6 illustrates representative changes after navigation-based reinforcement learning. In successful cases, ChapterNavRL tends to produce chapter titles that more clearly distinguish neighboring con-

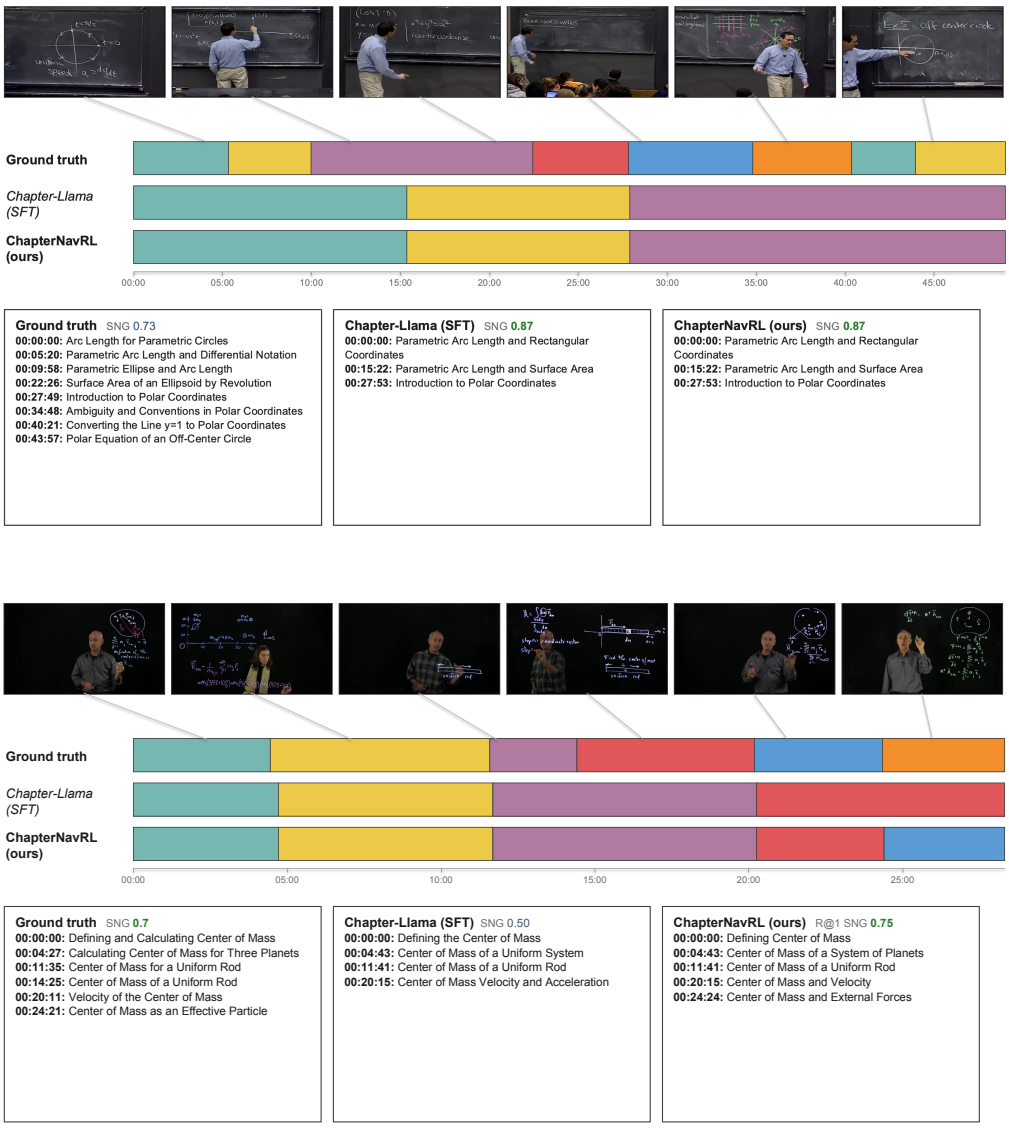

Figure 6: Representative comparison between SFT and ChapterNavRL. Navigation-based reinforcement learning can improve title discriminability and access efficiency while maintaining sufficient target coverage.

tent and temporal intervals that retain the required evidence while excluding unnecessary surrounding material.

These changes correspond to the components captured by the navigation evaluation. More informative titles improve title-based routing, while appropriately localized chapters reduce excess access cost without sacrificing target coverage. Importantly, the model is never given the evaluation query when generating the directory; the observed changes arise from optimizing reusable chapter structures across the collection of training-time navigation needs.