Table 1: Cause-and-Effect Analysis of Challenges in Ensuring High-Quality Data Annotation Requirements for AI-enabled Perception Systems (AIePS).

(Each entry includes the challenge, root causes, affected annotation phases, and empirically supported metadata such as severity, frequency, and interview ID references. This mapping enables traceable, systemic understanding of how annotation challenges emerge and propagate across the AIePS pipeline.)

Challenge	Causes	Consequences	Priority	Identified Phase	Impacted Phases	Severity	Frequency (how often a challenge was mentioned)	Empirical Support (Interview ID (explicitly mentioned or supported)
B1: Edge Case Coverage Gaps	Ambiguous guidelines, lack of domain expertise, limited training, sparse real-world data	System failures, rework costs, safety/legal risks, unreliable model behavior	High	Requirement Definition	Annotation Execution, Model Training, QA	Critical	High (16/19)	ID5, ID6, ID8
B1.1: Complexity in Annotating Edge Cases	Subjectivity, guideline interpretation, unclear metrics, insufficient standardization	Annotation inconsistency, degraded model reliability, compliance issues	High	Annotation Execution	Model Training, QA	High	High (15/19)	ID2, ID3, ID4, ID8, ID16, ID18,ID19
B1.2: Defining Requirements for Edge Cases	Ambiguity, lack of ground truth, reactive updates	Dataset blind spots, reduced annotation quality, error propagation	High	Requirement Definition	Annotation Execution, QA, Model Training	Critical	High (14/19)	ID2, ID3, ID5, ID16
B2: Ambiguity in Data Annotation Requirements	Unclear or evolving guidelines, lack of training, domain expertise, vague regulations	Inconsistent annotations, increased rework, degraded model accuracy	High	Requirement Definition	Annotation Execution, QA	Critical	High (17/19)	ID2, ID3, ID4, ID5, ID8, ID9, ID15, ID16, ID19
B3: Evolving Annotation Requirements	Lack of structured update workflows, reactive changes, tool mismatches	Inconsistencies across batches, model instability, increased rework	High	Requirement Evolution	Annotation Execution, QA, Model Training	High	High (16/19)	ID1, ID2, ID5, ID15, ID16, ID17
B4: Inconsistencies in Annotation Requirements	Lack of standardization, unclear processes, insufficient training and communication	Annotation variability, rework, reduced model reliability	High	Guideline Development, QA	Annotation Execution, Model Training	High	High (17/19)	ID2, ID3, ID4, ID5, ID15, ID19
B5: Resource Limitations	Budget constraints, time pressure, workforce shortage, poor tooling	Annotation errors, quality issues, scalability bottlenecks	High	Planning, Execution	All Phases	Critical	Very High (19/19)	ID1, ID2, ID3, ID5, ID6, ID8, ID9, ID12, ID15, ID16, ID17, ID19
B5.1: Strict Budgets	Low annotation funding, cost-saving pressures, limited tooling investment	Untrained workforce, reduced QA checks, tool limitations, higher error rates	High	Planning	Requirement Definition, Annotation Execution, QA	Critical	High (16/19)	ID3, ID5, ID6, ID8, ID9, ID17

B5.2: Limited Workforce and Scalability	Low pay, high turnover, custom onboarding, lack of automation	Quality inconsistencies, reduced annotation coverage, slower project progress	High	Execution	Annotation Execution, QA	High	High (15/19)	ID1, ID2, ID3, ID5, ID8, ID12, ID19
B5.3: Time Constraints Compromising Accuracy	Tight deadlines, speed-over-accuracy incentives, quota-driven shortcuts	Missed edge cases, elevated error rates, model performance degradation	High	Execution	Annotation Execution, Model Training	High	High (17/19)	ID3, ID5, ID6, ID7, ID9, ID12, ID16, ID19
B5.4: Limitations in Annotation Tools and Technology	Outdated tools, lack of automation, version conflicts, poor collaboration support	Annotation delays, inconsistent outputs, tool-induced data loss, model errors	High	Tooling Support, Execution	Annotation Execution, QA, Model Training	Critical	High (16/19)	ID1, ID3, ID6, ID9, ID15, ID17, ID19

Table 2. Criteria used to calculate Priority, Severity, and Frequency scores for practitioner recommendations and annotation challenges.

(The values are derived from cross-interview thematic analysis, considering the criticality of the challenge, its occurrence across interviews, and its impact on AI-enabled perception system (AIePS) performance.)

Dimension	Definition	How It's Measured	Heuristic Thresholds
Priority	Indicates the urgency and severity the identified challenge	Based on the severity of the challenge type, how often it is mentioned, and whether it is critical for safe or effective AIePS function	-High Priority: Must be addressed immediately due to critical impact or risk. -Medium Priority: Important but not urgent; should be resolved soon. -Low Priority: Can be delayed; limited impact if left unaddressed temporarily.
Severity	Reflects the negative impact of a challenge on data quality, model performance, or compliance	Inferred from interview narratives about the consequences of unresolved challenges (e.g., safety, legal, rework)	-High Severity: Causes major quality or consistency issues; may affect safety or compliance. -Moderate Severity: Impacts workflow efficiency, causes rework, but not critical. -Low Severity: Minor inconvenience; does not significantly affect outcomes.
Frequency	Represents how commonly a challenge was mentioned by interviewees	Counted via thematic coding across transcripts; can be expressed as ratios or categories	- High: Mentioned in ≥13 out of 19 interviews - Medium: Mentioned in 6–12 interviews - Low: Mentioned in ≤5 interviews