Ready Tensor Evaluation Rubric

Date: February 2025
Version: 1.0.0

Clear Purpose and Objectives

Description

- Evaluates whether the publication explicitly states its core purpose within the
  first paragraph or two."
- The purpose statement must clearly indicate what specific problem is being solved,
  what will be learned, or what will be demonstrated."
- Must appear in the abstract, tl;dr, introduction, or overview section and be
  immediately clear without requiring further reading."
- The key differentiator is an explicit, specific purpose statement near the top
  that lets readers immediately understand what the publication will deliver."

Scoring Logic

- Score 0: Purpose is unclear, appears too late, requires inference, or is too vague"
- Score 1: Explicit purpose statement appears in first paragraph/10 sentences and clearly
  states specific deliverables"

Positive Indicators

- States specific purpose in first paragraph
- Uses explicit purpose statement phrases (e.g., 'This paper

demonstrates...', 'In this guide, you will learn...')
- Clearly defines concrete deliverables or outcomes
- Lists specific skills or knowledge to be gained
- States exact problem being solved
- Defines precise scope of work
- Indicates specific contributions or solutions

Negative Indicators

- Purpose appears after first paragraph
- Requires reading multiple paragraphs to understand goal
- Lists multiple potential purposes
- Unclear what reader will gain
- Purpose scattered across document
- Ambiguous or general statements
- Purpose must be pieced together
- Only implies purpose without stating it
- Uses broad or non-specific language

Specific Objectives

Description

Assesses whether the publication lists specific and concrete objectives that will be addressed."

Scoring Logic

- Score 0: Missing specific objectives, or they are too vague
- Score 1: Contains specific objectives that clearly outline what will be achieved

Positive Indicators

- Lists specific things that will be covered/achieved

- Objectives are concrete and focused
- Objectives align with stated purpose
- Clear deliverables or outcomes specified

Negative Indicators

- Objectives are vague or too broad
- Missing explicit objectives
- Objectives don't align with stated purpose
- No clear outline of what will be achieved

Intended Audience/Use Case

Description

Evaluates whether the publication clearly identifies who it's for and how it benefits them."

Scoring Logic

- Score 0: No clear target audience or use case specified
- Score 1: Clear identification of target audience and intended use

Positive Indicators

- Specifies required background knowledge or other resources
- Identifies who will benefit from this content
- Explains how the content should be used or may benefit the audience

Negative Indicators

- No mention of target audience
- Unclear prerequisites or requirements
- Unclear why this study would matter to the audience

Target Audience Definition

Description

- Evaluates how well the publication identifies and describes the target audience
  for the tool, software package, dataset, or product, including user profiles, domains,
  and use cases."
- Should clearly communicate who would benefit from the solution and why, including
  technical requirements, usage context, and expected outcomes."
- Must specify both technical and practical prerequisites where applicable."

Scoring Logic

- Score 0: Target audience is unclear, too broad, or missing key details about user requirements
  and expected benefits"
- Score 1: Clear, specific definition of target users, their needs, and how the solution
  addresses those needs"

Positive Indicators

- Explicitly defines target domains or application areas
- Specifies technical skill requirements or prerequisites
- Describes typical user roles (e.g., developers, researchers, practitioners)
- Outlines specific use cases and scenarios
- Addresses scalability considerations
- Identifies relevant technical and non-technical stakeholders

Negative Indicators

- No clear audience definition
- Overly broad or vague target users
- Missing technical prerequisites
- Unclear fit between capabilities and user needs
- Fails to address different user perspectives

Specific Research Questions/Objectives

Description

Assesses whether the publication breaks down its purpose into specific, measurable
research questions or objectives that guide the investigation."

Scoring Logic

- Score 0: Missing specific questions/objectives, or they are too vague to guide research
- Score 1: Contains specific, measurable questions/objectives that clearly guide the work

Positive Indicators

- Lists numbered/bulleted specific questions or objectives
- Questions are concrete and focused (not overly broad)
- Questions are directly related to stated purpose
- Objectives are clearly measurable/assessable

Negative Indicators

- Questions are vague or too broad
- Objectives are not specific enough to guide investigation
- Missing explicit questions/objectives

- Questions don't align with stated purpose

Testability/Verifiability

Description

Assesses whether the research questions and hypotheses can be tested or verified
using the proposed approach. Research hypothesis must be falsifiable."

Scoring Logic

- Score 0: Questions/hypotheses cannot be adequately tested with proposed approach
- Score 1: Clear path to testing/verifying all questions/hypotheses

Positive Indicators

- Explains how each question will be answered
- Describes methods to verify hypotheses
- Links questions to specific tests/experiments
- Identifies data/evidence needed for verification

Negative Indicators

- No clear way to test stated hypotheses
- Questions cannot be answered with proposed methods
- Missing connection between questions and methodology
- Verification approach is unclear or missing

Literature Review Coverage & Currency

Description

Assesses the comprehensiveness and timeliness of literature review of similar works."

Scoring Logic

- Score 0: Literature review is missing, sparse or outdated; no references
- Score 1: Adequate and up-to-date coverage of relevant research papers

Positive Indicators

- Contains a references and/or relevant works section
- Cites recent works (2023-2024)
- Covers major developments in the field
- Demonstrates thorough research effort
- Balanced coverage of different approaches

Negative Indicators

- No or minimal references or citations
- Missing key relevant works
- Over-reliance on outdated sources
- Significant gaps in coverage
- Biased selection of references

Literature Review Critical Analysis

Description

Evaluates how well the publication analyzes and synthesizes existing work in literature."

Scoring Logic

- Score 0: No analysis of existing work;
  no citations or references to prior relevant works"
- Score 1: Thoughtful analysis and synthesis of existing work

Positive Indicators

- Contains a references and/or relevant works section
- Cites and analyzes strengths/weaknesses of existing research
- Compares different methodologies
- Synthesizes findings from multiple sources
- Provides insights about state of the art

Negative Indicators

- No or minimal references or citations
- Merely lists previous work without analysis
- No comparison between approaches
- Missing critical evaluation
- Superficial treatment of related work

Citation Relevance

Description

Evaluates whether the cited works are relevant and appropriately support the research context."

Scoring Logic

- Score 0: Missing citations or citations are irrelevant
- Score 1: Citations are relevant and support the work's context

Positive Indicators

- Contains a references or relevant works section
- Citations directly relate to the problem/methodology
- References support key claims and statements
- Citations include seminal works in the field
- References are properly attributed and contextualized

Negative Indicators

- Missing citations for key claims
- Citations are superficial or tangential
- References don't support the claims made
- Misrepresented or incorrectly cited works

Current State Gap Identification

Description

Assesses whether the publication clearly identifies gaps in existing work."

Scoring Logic

- Score 0: No gaps identified in existing work
- Score 1: Clear identification of gaps in existing research/work

Positive Indicators

- Contains a references and/or relevant works section
- Explicitly states limitations in current approaches from literature
- Identifies unaddressed problems/needs
- Points out areas needing improvement
- Shows understanding of field's current limitations

Negative Indicators

- No or minimal references or citations

- No discussion of limitations in existing work
- Fails to identify relevant gaps
- Unclear what problems remain unsolved
- Misidentifies or overstates gaps

Context Establishment

Description

Evaluates how well the publication establishes context for the topic covered."

Scoring Logic

- Score 0: No clear establishment of context
- Score 1: Clear establishment of context and justification for the topic of
  discussion"

Positive Indicators

- Builds clear foundation
- Shows why the work is necessary
- Connects past work to current objectives
- Demonstrates understanding of field's evolution

Negative Indicators

- Missing context for the work
- Weak or missing justification
- Fails to show work's relevance

Problem Definition

Description

Evaluates how well the publication defines and articulates the real-world problem
that motivated the AI/ML solution. This includes the problem's scope, impact,
and relevance to stakeholders."

Scoring Logic

- Score 0: Problem is unclear, poorly defined, or missing critical details about
  scope, impact, or stakeholder needs"
- Score 1: Clear, comprehensive definition of the problem with well-articulated
  scope, impact, and stakeholder needs"

Positive Indicators

- Clearly states the challenge or opportunity
- Quantifies the problem's impact (e.g., costs, time, resources, societal impact)
- Identifies key stakeholders and their needs
- Describes current solution/process limitations
- Provides relevant metrics or indicators
- Explains why AI/ML is an appropriate solution approach
- Defines concrete success criteria
- Outlines constraints and requirements
- Establishes the broader context (organizational, social, regulatory, etc.)

Negative Indicators

- Vague or generic problem statement
- Missing quantification of problem impact
- Unclear stakeholder needs or requirements
- No explanation of why current solutions are insufficient

- Lacks contextual background
- Focuses only on technical aspects without practical justification
- Missing success criteria or desired outcomes
- Fails to demonstrate real-world relevance

Methodology Explanation

Description

Evaluates whether the technical methodology is explained clearly and comprehensively,
allowing readers to understand the technical approach."

Scoring Logic

- Score 0: Missing or unclear methodology explanation
- Score 1: Clear methodology explanation with charts or tables where required.

Positive Indicators

- Clear explanation of technical approach
- Mathematical LaTex equations
- Logical flow of methodology
- Well-structured technical explanation
- Appropriate level of technical detail
- Diagrams or figures related to methodology are mentioned
- Supplementary materials for methodology is mentioned

Negative Indicators

- Vague or unclear explanations
- Missing key methodological steps
- Disorganized technical description
- Too high-level or too detailed

Step-by-Step Guidance Quality

Description

Evaluates how effectively the publication breaks down complex procedures into clear,
logical, and sequential steps that guide readers through the process. The steps should
build upon each other in a coherent progression, with each step providing sufficient
detail for completion before moving to the next."

Scoring Logic

- Score 0: Steps are unclear, missing critical details, out of sequence, or have significant
  gaps that would prevent readers from successfully following along"
- Score 1: Clear, well-structured sequence of steps that logically progress and contain
  sufficient detail for readers to successfully follow along"

Positive Indicators

- Clear numbering or ordering of steps
- Each step has a clear, specific objective
- Prerequisites or setup requirements are stated before relevant steps
- Steps build logically upon previous steps
- Appropriate level of detail for each step - neither too sparse nor overwhelming
- Critical decision points and options are clearly explained
- Complex procedures are broken down into manageable sub-steps
- Includes verification points to confirm successful completion
- References to previous steps when building upon them
- Clear transitions between major phases or sections

Negative Indicators

- Missing or unclear step sequence
- Steps require knowledge or actions not previously covered
- Inconsistent detail level across steps
- Large gaps between steps requiring readers to figure out intermediate steps
- Steps presented in illogical or inefficient order
- Critical steps missing or glossed over
- Unclear dependencies between steps
- No way to verify successful completion of steps
- Jumps between concepts without clear progression
- Assumes actions or knowledge without explanation

Assumptions Stated

Description

Evaluates whether technical assumptions are clearly stated and explained."

Scoring Logic

- Score 0: Technical assumptions not stated
- Score 1: Clear statement of technical assumptions

Positive Indicators

- Explicit listing of assumptions
- Explains reasoning behind assumptions
- Acknowledges limitations of assumptions
- States impact of assumptions

Negative Indicators

- Missing critical assumptions
- Unstated implicit assumptions
- Unclear assumption implications
- Unjustified assumptions

Solution Approach and Design Decisions

Description

Evaluates whether the overall solution approach and specific design decisions are
appropriate and well-justified. This includes explanation of methodology choice,
architectural decisions, and implementation choices. Common/standard approaches may
need less justification than novel or unconventional choices."

Scoring Logic

- Score 0: Solution approach and design choices not adequately explained or justified
- Score 1: Solution approach is either obviously appropriate OR well-justified if non-standard,
  AND key design decisions are clearly explained with alternatives considered"

Positive Indicators

- Uses and justifies appropriate overall approach
- Explains key architectural/design choices
- Discusses alternatives considered
- Analyzes trade-offs between options
- Justifies decisions based on requirements

- Documents why specific choices were preferred
- Acknowledges limitations of chosen solutions

Negative Indicators

- Unexplained solution approach
- Unclear design choices
- Missing alternative analysis
- No trade-off discussion
- Unjustified deviations from standard practices
- Ignores obvious alternatives
- Poor rationale for decisions

Experimental Protocol

Description

Assesses whether the publication outlines a clear, high-level approach for conducting the study."

Scoring Logic

- Score 0: No clear experimental protocol outlined
- Score 1: Clear, well-structured experimental protocol

Positive Indicators

- Outlines key phases/steps of the study
- Provides high-level workflow
- Shows logical sequence of investigation
- Explains overall approach

Negative Indicators

- Missing study workflow

- Unclear sequence of steps
- Disorganized approach
- Jumps directly to implementation details

Study Scope & Boundaries

Description

Evaluates whether the publication clearly defines the boundaries, assumptions, and
limitations of the study."

Scoring Logic

- Score 0: No clear definition of study scope or boundaries
- Score 1: Clear definition of what's included and excluded from the study

Positive Indicators

- Explicitly states what is in and out of scope
- Lists key assumptions made
- Defines constraints and limitations
- Clarifies boundary conditions

Negative Indicators

- Unclear study boundaries
- Missing important assumptions
- Undefined constraints
- Ambiguous scope

Evaluation Framework

Description

Assesses whether the publication defines a clear framework for evaluating results."

Scoring Logic

- Score 0: No clear evaluation framework
- Score 1: Clear evaluation framework defined

Positive Indicators

- Defines key metrics
- Specifies comparison baselines
- Outlines evaluation criteria
- Explains measurement approach

Negative Indicators

- Missing evaluation metrics
- Unclear success criteria
- No baseline comparisons
- Vague measurement approach

Validation Strategy

Description

Evaluates whether the publication outlines a clear approach to validating results."

Scoring Logic

- Score 0: No validation strategy outlined
- Score 1: Clear validation strategy defined

Positive Indicators

- Describes validation approach
- Outlines verification methods
- Considers robustness
- Includes quality checks

Negative Indicators

- Missing validation methods
- Unclear verification approach
- No quality assurance plan
- Ignored robustness considerations

Dataset Sources & Collection

Description

Evaluates whether dataset(s) used in the study are properly documented. For existing
datasets, proper citation and sourcing is required for each. For new datasets, the
collection methodology must be described. For benchmark studies or comparative analyses,
all datasets must be properly documented."

Scoring Logic

- Score 0: Missing or inadequate dataset source information
- Score 1: Source or collection information provided for all datasets

Positive Indicators

- {'For existing datasets': ['Clear citation for each dataset used', 'Links to original dataset sources', 'Proper attribution for all sources',

'Version/date of each dataset mentioned']}
- {'For new datasets': ['Collection methodology explained', 'Data gathering process documented', 'Sampling approach described', 'Collection timeline provided']}
- {'For multiple datasets': ['Clear organization of dataset information', 'Consistent documentation across all datasets', 'Relationship between datasets explained if relevant']}

Negative Indicators

- Missing sources for any dataset
- Unclear data origins
- No collection methodology for new data
- Incomplete sourcing information
- Missing attribution where required
- Inconsistent documentation across datasets

Dataset Description

Description

Assesses whether dataset(s) are comprehensively described, including their
characteristics, structure, content, and rationale for selection. For multiple datasets,
comparability and relationships should be clear."

Scoring Logic

- Score 0: Missing or inadequate dataset descriptions
- Score 1: Clear and complete description of all datasets

Positive Indicators

- Each dataset's size and scope described

- Data statistics provided for all datasets
- Class distributions (if applicable)
- Data quality characteristics
- Rationale for dataset selection explained

Negative Indicators

- Missing information for any dataset
- Unclear data structures
- No statistics provided
- Missing information on selection criteria

Data Requirements Specification

Description

- For implementations requiring data: evaluates whether the publication clearly
  specifies the data requirements needed."
- When data is needed, must define the expected data characteristics, format,
  and structure for the implementation to work."
- The key differentiator is clear specification of data needs that enables readers to prepare appropriate data for data-dependent implementations."
- Not applicable to implementations that don't require data (e.g., API setup guides)."

Scoring Logic

- Score 0: Publication requires data but lacks clear specification of data requirements"
- Score 1: Publication either clearly defines required data characteristics OR doesn't
  require data for implementation"

Positive Indicators

- Specifies required data fields/columns
- Describes expected data format
- States minimum data volume needed
- Explains data quality requirements
- Shows example data structure
- Details required data types
- Specifies data preprocessing needs
- Indicates data granularity
- States time period requirements
- Implementation doesn't require data

Negative Indicators

- Requires data but missing specifications
- Unclear data format when needed
- Vague about required data
- Assumes data availability
- No example data structure for data tasks
- Skips data requirements for data tasks
- Unclear preprocessing needs for data tasks

Dataset Selection or Creation

Description

Evaluates whether the rationale for dataset selection is explained, or for new
datasets, whether the creation methodology is properly documented."

Scoring Logic

- Score 0: Missing selection criteria or creation methodology
- Score 1: Clear explanation of selection or detailed creation

methodology

Positive Indicators

- {'For selected datasets': ['Selection criteria explained', 'Justification for choices', 'Suitability discussed']}
- {'For created datasets': ['Collection methodology detailed', 'Sampling approach explained', 'Creation process documented']}

Negative Indicators

- No explanation for dataset choice
- Missing creation methodology
- Unclear selection criteria
- Incomplete documentation of process

Datset procesing Methodology

Description

Evaluates whether data processing steps are clearly documented and justified.
This includes any preprocessing, missing data handling, anomalies handling,
and other data clean-up processing steps."

Scoring Logic

- Score 0: Missing or unclear processing steps
- Score 1: Properly documented processing methodology with justification

Positive Indicators

- Cleaning steps documented

- Data transformations explained
- Feature engineering described
- Handling of missing values
- Outlier handling
- Normalization/scaling approaches
- Data filtering criteria

Negative Indicators

- Missing preprocessing steps
- Unclear transformations
- Unexplained data handling
- Missing rationale for choices

Basic Dataset Stats

Description

Evaluates whether the publication provides clear documentation of fundamental
dataset properties"

Scoring Logic

- Score 0: Missing or unclear basic dataset characteristics
- Score 1: Comprehensive documentation of basic dataset characteristics

Positive Indicators

- Total number of samples/records
- Data formats and sizes
- Number of features/attributes
- Number of classes/labels if applicable
- Data types of each field

- File formats and organization
- Dataset splits if applicable (train/test/validation)
- Temporal coverage if applicable
- Spatial coverage if applicable

Negative Indicators

- Missing size information
- Unclear data formats
- Undocumented data types
- Missing split information when relevant
- Unclear coverage details when relevant

Implementation Details

Description

Assesses whether sufficient implementation details are provided with enough clarity.
Focuses on HOW the methodology was implemented.

Scoring Logic

- Score 0: Missing implementation related informatino
- Score 1: Clear implementation related information is provided

Positive Indicators

- Specific technical specifications
- Implementation steps detailed
- Code structure explained or code snippets shown
- Key implementation decisions documented
- Code repository or resources mentioned or linked
- Dataset documentation is linked or mentioned
- Supplementary materials for implementation are mentioned

Negative Indicators

- Missing crucial details of how the experiment was implemented
- Unclear implementation process
- Vague technical specifications
- Important steps omitted
- No mention of supplementary materials such as repository, other links, etc.

Parameters & Configuration

Description

Evaluates whether parameter choices and configuration settings are clearly specified
and justified where non-standard. Includes model hyperparameters, system configurations,
and any tuning methodology used."

Scoring Logic

- Score 0: Missing critical parameter specifications or using non-standard configurations
  without justification"
- Score 1: Clear parameter documentation with rationale and tuning methodology where relevant"

Positive Indicators

- Lists key parameters and their values
- Explains parameter selection process if non-standard
- Documents configuration settings
- Describes tuning methodology if relevant
- Provides reasoning for critical parameter choices

Negative Indicators

- Missing values for important parameters
- Unexplained non-standard parameter choices
- Unclear configuration settings
- No mention of tuning process where needed
- Using unusual settings without justification

Experimental Environment

Description

- Evaluates whether the computational environment and resources used for the work
  are clearly specified when relevant.
- This includes hardware specifications, software versions, and runtime environment details
  necessary for understanding the implementation context.
- This may not be applicable for some publications (e.g., theoretical work,
  high-level surveys, or conceptual overviews). In such cases, this criterion should
  receive a score of 0 with explanation noting it's not applicable."

Scoring Logic

- Score 0: Either not applicable for this type of publication, or missing critical environment
  specifications where they are needed
- Score 1: Clear specification of relevant environmental details and resource requirements

Positive Indicators

- Specifies hardware used (CPU, GPU, memory if relevant)

- States computational resources required
- Provides relevant software versions
- Documents runtime environment
- Includes execution times or resource usage where relevant

Negative Indicators

- Missing hardware specifications where performance is discussed
- Unclear computational requirements
- Unspecified runtime environment where relevant
- No indication of resource needs
- Missing crucial version information

Tools, Frameworks, & Services

Description

- Documents the key tools, frameworks, 3rd party services used in the implementation when relevant.
- Standard/mainstream tools (e.g., PyTorch, TensorFlow, HuggingFace) need only be mentioned without justification.
- Non-standard or custom tools should be explained.
- Third party tools and services should be explained.
- This may not be applicable for some publications (e.g., theoretical work, surveys) and should receive a score of 0 with explanation noting it's not applicable.

Scoring Logic

- Score 0: Either not applicable for this type of publication, or missing critical tool/framework information where needed.
- Score 1: Clear specification of tools/frameworks used. If non-standard choices made, they are

properly justified.

Positive Indicators

- Lists main tools and frameworks used
- Justifies use of non-standard or custom tools
- Specifies versions for reproducibility where relevant
- Explains custom implementations when used
- Documents tool integration if complex

Negative Indicators

- Missing critical tool/framework information
- Unexplained use of non-standard tools
- Unclear custom implementation details
- Missing version information where needed
- No explanation for avoiding standard tools

Implementation Considerations

Description

Evaluates coverage of practical aspects of implementing or applying the model, concept,
app, or tool described in the publication.

Scoring Logic

- Score 0: Missing or inadequate coverage of implementation considerations
- Score 1: Clear discussion of practical implementation aspects

Positive Indicators

- Discussion of common challenges

- Resource requirements or constraints
- Scalability considerations
- Performance implications
- Integration aspects
- Best practices for implementation

Negative Indicators

- No discussion of practical challenges
- Missing important constraints
- Overlooking real-world limitations
- Ignoring implementation complexity
- No guidance on implementation

Deployment Considerations

Description

Evaluates whether the publication adequately discusses deployment requirements,
considerations, and challenges for implementing the solution in a production
environment. This includes either actual deployment details if deployed, or thorough
analysis of deployment requirements if proposed."

Scoring Logic

- Score 0: Missing or inadequate discussion of deployment considerations
- Score 1: Clear discussion of deployment requirements, considerations, or actual deployment
  details if implemented"

Positive Indicators

- Discusses infrastructure requirements
- Addresses resource needs (compute, memory, storage)
- Considers system dependencies
- Explains integration requirements
- Discusses scalability considerations
- Addresses security requirements
- Mentions monitoring needs
- Considers performance requirements
- Details environment setup
- Discusses potential challenges

Negative Indicators

- No mention of deployment requirements
- Missing infrastructure considerations
- Ignores resource requirements
- Overlooks integration needs
- Fails to address practical deployment challenges
- No discussion of production environment needs

Monitoring and Maintenance Considerations

Description

Evaluates whether the publication discusses how to monitor the solution's performance
and maintain its effectiveness over time. This includes monitoring strategies, maintenance
requirements, and operational considerations for keeping the solution running optimally."

Scoring Logic

- Score 0: Missing or inadequate discussion of monitoring and maintenance needs
- Score 1: Clear discussion of monitoring approaches and maintenance requirements for
  keeping the solution operational and effective"

Positive Indicators

- Discusses key metrics to monitor
- Describes logging requirements
- Addresses performance monitoring
- Explains model/system health checks
- Discusses data drift detection
- Outlines maintenance schedules
- Describes update procedures
- Addresses error handling
- Explains troubleshooting approaches
- Considers retraining requirements

Negative Indicators

- No mention of monitoring needs
- Missing maintenance considerations
- Overlooks system health tracking
- Ignores performance degradation
- No discussion of updates/maintenance
- Fails to address operational issues

Performance Metrics Analysis

Description

Evaluates whether appropriate performance metrics are used and properly analyzed
to demonstrate the success or effectiveness of the work."

Scoring Logic

- Score 0: Missing or inappropriate performance metrics
- Score 1: Relevant and comprehensive performance analysis with multiple relevant metrics"

Positive Indicators

- Appropriate metrics for the task
- Clear presentation of results
- Metrics properly explained
- Results properly quantified
- Error margins/variance reported

Negative Indicators

- Missing crucial metrics
- Inappropriate metric choice
- Unclear measurement approach
- Incomplete results reporting
- Missing error analysis

Comparative Analysis

Description

Assesses whether results are properly compared against relevant baselines or
state-of-the-art alternatives. At least 4 or 5 alternatives are compared with."

Scoring Logic

- Score 0: No comparison with baselines or state-of-the-art alternatives; or,

not sufficient comparisons (e.g. 4 or 5 alternatives)"
- Score 1: Clear comparison with appropriate baselines/alternatives

Positive Indicators

- Relevant baselines identified
- SOTA comparisons and benchmarks
- Fair comparison methodology
- Performance differences analyzed
- Trade-offs discussed
- Consistent comparison metrics

Negative Indicators

- Missing relevant comparisons
- Unfair comparison methodology
- Incomplete analysis
- Inconsistent metrics

Statistical Analysis

Description

Evaluates whether appropriate statistical methods are used to validate results."

Scoring Logic

- Score 0: Missing or inappropriate statistical analysis
- Score 1: Appropriate statistical analysis of results

Positive Indicators

- Proper statistical tests used
- Significance levels reported

- Confidence intervals provided
- Statistical assumptions verified
- Robustness checks performed

Negative Indicators

- Missing statistical validation
- Inappropriate statistical methods
- Unclear significance levels
- Statistical assumptions violated

Key Results

Description

Evaluates whether the main results and outcomes of the research are clearly
presented in an understandable way."

Scoring Logic

- Score 0: Key results missing or unclear
- Score 1: Clear presentation of main results

Positive Indicators

- Clearly states what the research found
- Presents important outcomes or measurements
- Uses appropriate numbers/metrics where relevant
- Includes relevant comparisons if applicable
- Uses clear visualizations if needed
- Makes results understandable without oversimplifying
- Maintains accuracy while being accessible

Negative Indicators

- Missing key results
- Results presented too technically
- Important findings omitted
- Results too vague or oversimplified
- Unclear what was actually found
- Missing important measurements
- Results presented without context

Results Interpretation

Description

Assesses whether results are properly interpreted and their implications explained."

Scoring Logic

- Score 0: Missing or inadequate interpretation of results
- Score 1: Clear and well-reasoned interpretation

Positive Indicators

- Clear explanation of findings
- Results contextualized
- Implications discussed
- Insights highlighted
- Unexpected results addressed

Negative Indicators

- Missing interpretation
- Superficial analysis
- Overreaching conclusions
- Ignored anomalies

Solution Impact Assessment

Description

Evaluates how well the publication quantifies and demonstrates the real-world
impact and value created by implementing the AI/ML solution. This includes measuring
improvements in organizational metrics (cost savings, efficiency gains, productivity),
user-centered metrics (satisfaction, adoption, time saved), and where applicable,
broader impacts (environmental, societal benefits). The focus is on concrete
outcomes and value creation, not technical performance measures."

Scoring Logic

- Score 0: Impact is not quantified, poorly measured, or relies solely on technical metrics
  without connecting to real-world value"
- Score 1: Clear quantification of impact with specific, measurable improvements in relevant
  metrics such as cost savings, efficiency gains, user satisfaction, or societal benefits"

Positive Indicators

- Quantifies improvements in key performance indicators
- Presents clear before-and-after comparisons
- Measures cost savings or efficiency gains
- Demonstrates time or resource savings
- Shows improvement in user/stakeholder satisfaction
- Quantifies environmental impact where applicable
- Measures societal benefits where relevant
- Includes return on investment calculations

- Reports productivity improvements
- Shows reduction in errors or risks
- Demonstrates scalability of impact
- Quantifies both direct and indirect benefits
- Provides concrete metrics such as cost reduction percentage, process time reduction, satisfaction scores
- Shows impact on resource utilization and allocation
- Measures compliance or risk reduction benefits

Negative Indicators

- Focuses only on technical metrics (accuracy, precision, etc.)
- Vague or qualitative statements about impact
- Missing baseline comparisons
- No concrete numbers or measurements
- Fails to connect improvements to stakeholder value
- Omits important impact dimensions
- No long-term impact assessment
- Missing cost-benefit analysis
- Unrealistic or unsubstantiated claims
- Lacks quantifiable evidence of improvements
- No measurement of user or stakeholder benefits

Constraints, Boundaries, and Limitations

Description

- Evaluates whether the publication clearly defines when and where the work is
  applicable (boundaries), what constrains its effectiveness (constraints), and
  what its shortcomings are (limitations)."
- Must identify specific circumstances where the solution works best,

where it
struggles, and where it fails or is unsuitable."
- Should discuss both technical and practical aspects that limit or constrain
the work's applicability or effectiveness."
- The key differentiator is transparent communication about the bounds of
applicability and effectiveness, along with known limitations."

Scoring Logic

- Score 0: Publication does not adequately discuss constraints, boundaries of applicability,
and limitations"
- Score 1: Publication clearly defines where/when the work is effective, what constrains it,
and what its limitations are"

Positive Indicators

- Specifies where/when solution works best
- Identifies where solution struggles
- States performance boundaries
- Describes scaling constraints
- Acknowledges known shortcomings
- Discusses performance trade-offs
- Identifies failure cases
- Explains solution constraints
- Details accuracy limitations
- Discusses edge cases
- Acknowledges inherent limitations

Negative Indicators

- No discussion of boundaries
- Omits known limitations
- Unclear about applicability

- Overstates capabilities
- Ignores known issues
- Vague about constraints
- Fails to define limits
- Unrealistic claims
- Hides known problems
- Only lists basic requirements

Key Findings

Description

Evaluates whether the main findings and contributions of the work are clearly
summarized and their significance explained."

Scoring Logic

- Score 0: Missing or inadequate summary of findings
- Score 1: Appropriate and well-structured summary of key findings

Positive Indicators

- Clear summary of main results
- Key contributions highlighted
- Findings properly contextualized
- Major insights emphasized
- Connection to original objectives shown

Negative Indicators

- Missing important findings
- Unclear main contributions
- Disconnected from objectives
- No clear takeaways

Significance and Implications of Work

Description

Assesses whether the broader significance and implications of the work are properly
discussed."

Scoring Logic

- Score 0: No discussion of implications
- Score 1: Clear discussion of work's implications

Positive Indicators

- Why this matters explained
- Practical implications explained
- Solves important problems
- Theoretical implications discussed
- Impact on field addressed
- Real-world applications noted
- Significance of contributions explained
- Valuable for decision making
- Impacts future research

Negative Indicators

- Missing implications discussion
- Unclear significance
- Overstated impact
- Disconnected from findings
- Limited consideration of broader context
- Limited practical value

- Minor theoretical impact
- Unclear importance
- Limited applicability
- Minimal influence on field

Features and Benefits Analysis

Description

- Evaluates the clarity and completeness of feature descriptions and their
  corresponding benefits to users."
- Should connect technical capabilities to practical outcomes and user value."
- Must include both functional features and their practical applications."

Scoring Logic

- Score 0: Features are poorly described, benefits are unclear, or connection between
  features and value is weak"
- Score 1: Clear description of key features with well-articulated benefits and value
  proposition"

Positive Indicators

- Comprehensive list of core features
- Clear mapping between features and benefits
- Quantified performance metrics where applicable
- Real-world application examples
- Integration capabilities and requirements
- Technical specifications and limitations
- Usage costs or resource requirements (if applicable)
- Installation and deployment requirements

- API documentation or interface descriptions

Negative Indicators

- Missing critical feature information
- Features listed without clear benefits
- Vague or promotional descriptions
- Lack of technical specifications
- Missing integration details
- Unclear value proposition

Competitive Differentiation

Description

- Evaluates how effectively the publication demonstrates the solution's unique value
  proposition and advantages compared to alternatives."
- Should highlight distinctive features, capabilities, or approaches that set the
  solution apart from existing alternatives."
- Must provide concrete evidence or examples of advantages."

Scoring Logic

- Score 0: Advantages are unclear, unsupported, or fail to differentiate from alternatives"
- Score 1: Clear, well-supported demonstration of unique value proposition and advantages
  over alternatives"

Positive Indicators

- Specific comparison points with alternative solutions
- Quantified performance or efficiency advantages

- Unique technical capabilities or approaches
- Novel feature combinations or integrations
- Specific use-case advantages
- User testimonials or adoption examples
- Independent benchmarks or evaluations
- Open source benefits (if applicable)
- Licensing advantages

Negative Indicators

- No clear differentiation from alternatives
- Unsupported claims of superiority
- Vague or generic advantages
- Missing comparison with alternatives
- Lack of specific examples or evidence
- Unclear positioning in the ecosystem

Future Directions

Description

Evaluates whether meaningful future work and research directions are identified."

Scoring Logic

- Score 0: No discussion of future directions
- Score 1: Clear and meaningful future directions identified

Positive Indicators

- Specific future work outlined
- Research gaps identified
- Potential improvements suggested
- New research questions posed

- Extension possibilities discussed

Negative Indicators

- Missing future directions
- Vague suggestions
- Obvious/trivial extensions
- No connection to limitations
- No clear path forward

Originality of Work

Description

Evaluates whether the work presents an original contribution, meaning work
that hasn't been done before. This includes novel analyses, comprehensive
comparisons, new methodologies, or new implementations.
Example 1: A benchmark study comparing existing methods would score 1 here,
if it's the first benchmark study even if the methods have existed before.
Example 2: A tool developed by integrating existing tools in new ways would
be original work; or an existing tool modified to work in a new industry
or domain would also be considered original."

Scoring Logic

- Score 0: Repeats or replicates existing work
- Score 1: Clear original contribution not previously published

Positive Indicators

- First study, model, technique, tool or app of its kind

- Novel analysis or comparison
- Unique combination of existing methods
- New application with substantial analysis
- Original experimental design

Negative Indicators

- Replicates existing studies
- Similar work already published
- Minor variations of version update of existing work
- Standard application without new analysis

Innovation in Methods/Approaches

Description

Evaluates whether the authors created new methods, algorithms, or applications.
This specifically looks for technical innovation, not just original analysis.
Example: A benchmark study comparing existing methods would score 0 here, even
if original and valuable.
Example 2: An existing tool modified to work in a new domain would not be
considered an innovation."

Scoring Logic

- Score 0: No new methods/approaches created
- Score 1: Created new method or approach

Positive Indicators

- New algorithm created
- Novel method developed

- New technical approach introduced
- Original framework created
- Innovative solution designed

Negative Indicators

- Only uses existing methods
- Applies standard approaches
- Combines existing techniques
- Parameter modifications only
- Implementation variations only

Advancement of Knowledge or Practice

Description

Evaluates how the work advances knowledge or practice, whether through original analysis or innovative methods or implementation.

Scoring Logic

- Score 0: No clear advancement of knowledge or practice
- Score 1: Clear knowledge or practice advancement with evidence

Positive Indicators

- New insights provided
- Knowledge gaps filled
- Performance differences quantified
- Limitations identified
- Better understanding enabled

Negative Indicators

- No new insights
- Minor or trivial findings
- Unclear conclusions
- Weak evidence
- Limited analysis

Code & Dependencies

Description

Evaluates whether code is available and dependencies are properly documented
for reproduction."

Scoring Logic

- Score 0: Code not linked or mentioned, or not applicable
- Score 1: Code repository is linked or mentioned

Positive Indicators

- Code repository provided
- Dependencies listed
- Installation instructions clear
- Package requirements specified
- Setup scripts available

Negative Indicators

- Missing code
- Unclear dependencies
- Poor documentation
- Installation gaps
- Missing requirements

Data Source and Collection

Description

Evaluates whether the publication clearly describes where the data comes from
and the strategy for data collection or generation. This criterion only applies
if the publication involved sourcing and creation of the data by authors."

Scoring Logic

- Score 0: Either Data sources or collection strategy not clearly described; or
  the authors didnt themselves create the dataset involved in the publication."
- Score 1: Clear description of data sources and systematic collection/generation strategy"

Positive Indicators

- Clearly identifies all data sources
- Explains why these sources were chosen
- Details the collection or generation strategy
- Describes any sampling approaches used
- Documents any data purchase or licensing arrangements
- API documentation and credentials process
- Data source URLs and access methods
- Database connection details
- Web scraping targets and permissions
- Data purchase/licensing instructions
- Generation scripts for synthetic data
- Annotation guidelines for manual data

Negative Indicators

- Unclear or incomplete source information
- Missing rationale for source selection
- Ad-hoc or poorly defined collection strategy
- Undocumented sampling decisions
- Missing source information
- Unclear access methods
- Undocumented permissions
- Missing generation scripts
- Incomplete annotation guidelines

Data Inclusion and Filtering Criteria

Description

Assesses whether the publication defines clear criteria for what data is included or excluded from the dataset"

Scoring Logic

- Score 0: Missing or unclear inclusion/exclusion criteria
- Score 1: Well-defined criteria for data inclusion and exclusion

Positive Indicators

- Clear rules for data inclusion
- Explicit exclusion criteria
- Rationale for filtering decisions
- Documentation of edge cases
- Quantification of filtered data

Negative Indicators

- No clear inclusion/exclusion rules

- Missing justification for criteria
- Inconsistent application of criteria
- Undocumented filtering decisions

Dataset Creation Quality Control Methodology

Description

Evaluates the systematic approach to ensuring data quality during collection, generation, and processing"

Scoring Logic

- Score 0: Missing or inadequate quality control methodology
- Score 1: Well-defined quality control process with clear validation steps

Positive Indicators

- Defined quality control process
- Clear validation checkpoints
- Error detection methods
- Quality assurance procedures
- Documentation of quality metrics

Negative Indicators

- No quality control process
- Missing validation steps
- Unclear quality standards
- Ad-hoc quality checks

Dataset Bias and Representation Consideration

Description

Assesses whether potential biases in data collection/generation are identified
and addressed. For synthetic or naturally bias-free datasets, clear documentation
of why bias is not a concern is sufficient."

Scoring Logic

- Score 0: No consideration of potential biases or representation issues where relevant"
- Score 1: Either:
- Thorough analysis of potential biases and representation considerations
  OR
- Clear explanation of why bias is not a concern (e.g., synthetic data with
  controlled generation, complete population coverage, or bias-irrelevant
  technical datasets)"

Positive Indicators

- Identifies potential sources of bias where relevant
- Analyzes data representation where applicable
- Documents demographic distribution if relevant
- Addresses sampling bias where applicable
- Discusses limitations and skews if present
- Clearly explains why bias is not a concern for bias-irrelevant datasets
- Documents the controlled generation process for synthetic data
- Demonstrates comprehensive coverage for complete population

datasets

Negative Indicators

- No bias analysis for potentially biased data sources
- Missing representation assessment where relevant
- Unaddressed sampling issues in sampled datasets
- Ignored demographic skews in human-related data
- No explanation of bias-free nature for synthetic/technical datasets
- Assumes bias-free without justification

Statistical Characteristics

Description

Assesses whether the publication provides comprehensive statistical information
about the dataset"

Scoring Logic

- Score 0: Missing or inadequate statistical analysis
- Score 1: Either:
- Comprehensive statistical analysis of dataset characteristics
  OR
- Clear explanation of why certain statistics are not applicable"

Positive Indicators

- Class/label distributions if applicable
- Feature distributions
- Summary statistics for numerical features
- Missing value statistics
- Outlier analysis
- Correlation analysis if relevant

- Time series characteristics if applicable
- Demographic distributions if applicable

Negative Indicators

- Missing distribution information
- Unclear statistical summaries
- Undocumented missing values
- Missing relevant analyses
- No explanation for omitted statistics

Dataset Quality Metrics and Indicators

Description

Evaluates whether the publication provides clear metrics and indicators of data quality"

Scoring Logic

- Score 0: Missing or inadequate quality metrics
- Score 1: Comprehensive quality metrics with clear assessment methodology

Positive Indicators

- Data completeness metrics
- Consistency checks
- Noise level assessment
- Label quality metrics if applicable
- Inter-annotator agreement if applicable
- Data integrity checks
- Source reliability assessment

- Quality control results

Negative Indicators

- Missing quality assessments
- Unclear quality metrics
- Undocumented quality issues
- Missing validation results
- Poor quality control documentation

State-of-the-Art Comparisons

Description

Evaluates whether the study includes relevant state-of-the-art methods from
recent literature for comparison. Must contain at least 4 or 5 other top methods
for comparison"

Scoring Logic

- Score 0: Missing key SOTA methods or using outdated comparisons
- Score 1: Includes relevant current SOTA methods - at least 4 or 5 top methods"

Positive Indicators

- Recent SOTA methods included
- Current top performers covered
- Up-to-date implementations used
- Latest variants considered
- Key competing approaches included

Negative Indicators

- Missing important SOTA methods
- Using outdated versions
- Ignoring recent advances
- Incomplete coverage of field
- Only older methods compared

Benchmarking Method Selection Justification

Description

Evaluates whether the choice of methods, models, or tools for comparison is
well-justified and reasonable for the study's objectives."

Scoring Logic

- Score 0: No clear justification for method selection
- Score 1: Clear rationale for methods chosen and excluded

Positive Indicators

- Selection criteria explained
- Exclusions justified
- Relevance demonstrated
- Alternative considerations discussed
- Method scope appropriate

Negative Indicators

- Arbitrary selection
- Unexplained omissions
- Biased choices
- Missing key alternatives

- Poor scope definition

Fair Comparison Setup

Description

Assesses whether all methods are compared under fair and consistent conditions."

Scoring Logic

- Score 0: Unfair or inconsistent comparison conditions
- Score 1: Demonstrated fairness in comparisons

Positive Indicators

- Consistent data preprocessing
- Equal computational resources
- Proper hyperparameter tuning
- Same evaluation conditions
- Controlled testing environment

Negative Indicators

- Inconsistent conditions
- Unfair resource allocation
- Poor parameter tuning
- Biased evaluation setup
- Uncontrolled variables

Benchmarking Evaluation Rigor

Description

Evaluates whether the comparison uses appropriate metrics and statistical analysis."

Scoring Logic

- Score 0: Poor or insufficient evaluation methodology
- Score 1: Appropriate evaluation methodology with proper metrics

Positive Indicators

- Multiple relevant metrics
- Statistical significance tests
- Error margins reported
- Variance analysis included
- Proper statistical methodology

Negative Indicators

- Single/inappropriate metrics
- Missing statistical tests
- No error analysis
- Poor statistical methodology
- Incomplete results reporting

Purpose-Aligned Topic Coverage

Description

Evaluates whether the publication covers all topics and concepts necessary to fulfill
its stated purpose, goals, or learning objectives. Coverage should be complete relative
to what was promised, rather than exhaustive of the general topic area."

Scoring Logic

- Score 0: Missing topics necessary to fulfill stated purpose, or includes substantial
  unrelated content"
- Score 1: Completely covers all topics needed to fulfill stated purpose, with appropriate
  focus and depth"

Positive Indicators

- Each stated objective or goal has corresponding content coverage
- Topics directly support the publication's stated purpose
- Content scope matches what was promised
- Prerequisites stated in objectives are properly covered
- For tutorials/guides: All steps needed to achieve stated learning outcomes
- For technical reports: All components needed to understand and replicate work
- For case studies: All aspects needed to understand context, solution, and outcomes
- For research papers: All elements needed to validate and understand findings
- Logical progression of topics that builds toward stated goals
- Appropriate depth for each topic based on its importance to stated objectives
- Clear connections between covered topics and stated purposes
- Balanced coverage aligned with topic's importance to objectives

Negative Indicators

- Stated objectives not fully addressed in content
- Missing topics necessary to achieve stated goals
- Gaps in prerequisite knowledge promised to be covered
- Important aspects mentioned in purpose statement left unexplained
- Excessive coverage of topics not related to stated purpose
- Disproportionate focus on secondary aspects
- Missing crucial steps or components needed for stated outcomes
- Coverage doesn't support claimed deliverables

- Topics presented without clear connection to objectives
- Critical dependencies for stated goals left unexplained
- Insufficient depth in areas central to stated purpose
- Imbalanced coverage relative to topics' importance to goals

---

Clear Prerequisites and Requirements

Description

Evaluates whether the publication clearly states what readers need to have (tools,
environment, software) or need to know (technical knowledge, concepts) before they
can effectively use or understand the content. Most relevant for educational content
like tutorials, guides, and technical implementations, but can also apply to technical deep dives and implementation reports."

Scoring Logic

- Score 0: Missing or unclear prerequisites that readers need for successful understanding
  or implementation"
- Score 1: Clear, complete statement of all necessary prerequisites and requirements

Positive Indicators

- Clearly states required background knowledge or expertise level
- Lists specific concepts readers should be familiar with
- Indicates expected proficiency level in relevant technologies
- Mentions if specific math/statistics knowledge is needed
- Lists required software versions and dependencies

- Specifies hardware requirements if applicable
- Details environment setup needs
- Mentions necessary access rights or accounts
- Prerequisites section appears early in the content
- Distinguishes between required and optional prerequisites
- Groups requirements logically (knowledge vs. tools)
- Links to resources for acquiring prerequisites when applicable

Negative Indicators

- No mention of required background knowledge
- Missing critical software requirements
- Unclear minimum expertise level needed
- Hidden requirements that only become apparent later
- Vague statements about required knowledge
- Ambiguous version requirements
- Unclear distinction between must-have and nice-to-have
- Prerequisites scattered throughout content instead of clearly stated upfront
- Assumes knowledge not listed in prerequisites
- Uses tools or technologies not mentioned in requirements
- References concepts without checking if they're in prerequisites
- Inconsistency between stated prerequisites and actual content complexity

---

Appropriate Technical Depth

Description

Assesses whether the technical content matches the expected depth for the intended
audience and publication type. For technical audiences, evaluates if it provides
sufficient depth. For general audiences, evaluates if it maintains accessibility

while being technically sound."

Scoring Logic

- Score 0: Technical depth misaligned with audience needs (either too shallow or too complex)"
- Score 1: Technical depth appropriately matched to audience and publication type"

Positive Indicators

- When targeting technical audiences:
- Comprehensive technical explanations
- Mathematical foundations and formal notations where relevant
- Implementation details and architectural considerations
- Analysis of edge cases and limitations
- Performance characteristics and trade-offs
- When targeting intermediate audiences:
- Clear explanations of core concepts
- Simplified mathematical notation with explanations
- Basic implementation guidance
- Common pitfalls and best practices
- Practical examples and use cases
- When targeting general audiences:
- High-level explanations without jargon
- Visual explanations of complex concepts
- Practical implications and business value
- Real-world analogies and examples
- Clear connection to practical applications

Negative Indicators

- Technical depth doesn't match stated audience level
- Unexplained technical jargon for non-technical audiences
- Oversimplified explanations for technical audiences
- Missing critical details expected by the target audience
- Sudden jumps in complexity level

- Inconsistent technical depth throughout
- Using advanced concepts without proper introduction
- Assuming knowledge not stated in prerequisites
- Important technical details missing for technical audiences
- Too much technical detail for general audiences
- Mathematical notation without proper explanation for non-technical audiences
- Missing practical context for technical concepts

Code Usage Appropriateness

Description

Assesses whether code examples, when present, are used judiciously and add value
to the explanation. If the publication type or topic doesn't require code examples,
then absence of code is appropriate and should score positively."

Scoring Logic

- Score 0: Unnecessary code dumps or excessive implementation details where code is used
- Score 1: Either: appropriate absence of code OR focused code examples that directly support explanations

Positive Indicators

- Appropriate absence of code in non-technical content
- Code examples directly illustrate specific concepts
- Minimal, focused snippets rather than complete implementations
- Clear purpose for each code example
- Balance between code and explanation

Negative Indicators

- Unnecessary code where concepts could be explained without it
- Excessive implementation details
- Code dumps without proper explanation
- Unnecessary boilerplate code
- Code overshadowing conceptual content

Code Clarity and Presentation

Description

When code examples are present, evaluates whether they are well-written, properly
formatted and integrated with the surrounding content. If the publication contains
no code examples, this criterion is considered satisfied by default."

Scoring Logic

- Score 0: Code is present but poorly presented or formatted
- Score 1: Either: no code present OR code is clear and well-formatted

Positive Indicators

- No code present (automatically satisfies criterion)
- {'When code is present': ['Clear variable and function names', 'Consistent formatting and style', 'Code broken into digestible parts', 'Appropriate abstraction level', 'Good visual presentation']}

Negative Indicators

- {'When code is present': ['Unclear or confusing code structure', 'Inconsistent formatting', 'Poor organization', 'Inappropriate abstraction level', 'Poor visual presentation']}

Code Explanation Quality

Description

When code snippets are present, evaluates how well they are explained and
contextualized within the content. If the publication contains no code snippets,
this criterion is considered satisfied by default."

Scoring Logic

- Score 0: Code is present but poorly explained or contextualized
- Score 1: Either: no code present OR code is well explained and contextualized

Positive Indicators

- No code present (automatically satisfies criterion)
- {'When code is present': ['Each significant code snippet is explained', 'Key lines or sections are highlighted', 'Implementation choices are justified', 'Important details are discussed', 'Connects code to concepts']}

Negative Indicators

- {'When code is present': ['Missing explanations for snippets', 'Superficial explanations only', 'Complex parts left unexplained', 'No context for implementation choices', 'Disconnect from main content']}

Real-World Applications

Description

Assesses whether the publication clearly explains the practical significance,
real-world relevance, and potential applications of the topic. This shows readers
why the content matters and how it can be applied in practice."

Scoring Logic

- Score 0: Missing or unclear practical relevance and applications
- Score 1: Clear explanation of practical importance with concrete applications

Positive Indicators

- Clear explanation of why the topic matters in practice
- Specific industry or domain applications identified
- Business or technical impact explained
- Implementation considerations discussed
- Different application contexts provided
- Real-world constraints and limitations addressed

Negative Indicators

- No connection to practical applications
- Purely theoretical discussion
- Missing real-world context
- Too abstract or academic
- No mention of practical value
- Ignores real-world constraints

Limitations and Trade-offs

Description

Assesses whether the content discusses practical limitations, trade-offs,

and
potential pitfalls in real-world applications.

Scoring Logic

- Score 0: Missing or inadequate coverage of limitations and trade-offs
- Score 1: Clear discussion of practical limitations and trade-offs

Positive Indicators

- Discussion of limitations and constraints
- Clear trade-off analysis
- Common pitfalls and how to avoid them
- When not to use certain approaches
- Alternative approaches for different scenarios
- Cost-benefit considerations

Negative Indicators

- No discussion of limitations
- Missing important trade-offs
- Overselling capabilities
- No mention of pitfalls
- Ignoring alternative approaches
- Missing cost-benefit analysis

Supporting Examples

Description

Evaluates whether educational content (tutorials, guides, blogs, technical deep dives)
includes concrete and contemporary examples to illustrate concepts and enhance understanding.
Examples should help readers better grasp the material through practical

demonstration."

Scoring Logic

- Score 0: Missing examples or examples that don't effectively illustrate concepts,
  or outdated exammples."
- Score 1: Includes clear, relevant examples that enhance understanding

Positive Indicators

- Provides specific examples to illustrate key concepts
- Examples progress from simple to complex
- Examples demonstrate practical usage
- Examples cover different scenarios or use cases
- Examples include common pitfalls or edge cases
- Code examples where appropriate
- Step-by-step worked examples
- Examples match the audience's level

Negative Indicators

- Concepts presented without supporting examples
- Examples too complex for the target audience
- Examples don't clearly illustrate the concepts
- Missing examples for key points
- Examples without proper explanation
- Disconnected or irrelevant examples
- Only trivial examples provided

Industry Insights

Description

Evaluates inclusion of industry trends, statistics, or patterns observed in practice.

Scoring Logic

- Score 0: Missing or minimal industry context
- Score 1: Includes relevant industry trends and data

Positive Indicators

- Current industry trends
- Usage statistics or adoption rates
- Market observations
- Industry benchmarks
- Technology adoption patterns
- Industry-specific challenges

Negative Indicators

- No industry context
- Outdated information
- Missing relevant trends
- No statistical backing
- Generic observations only

Success/Failure Stories

Description

Assesses whether specific success or failure stories are shared to illustrate
outcomes and lessons learned.

Scoring Logic

- Score 0: Missing or minimal real outcome stories
- Score 1: Includes specific success/failure stories with lessons learned

Positive Indicators

- Specific project outcomes
- Real implementation results
- Detailed failure analysis
- Success story breakdowns
- Lessons learned from failures
- Unexpected challenges faced

Negative Indicators

- No specific outcomes shared
- Missing failure analysis
- Generic success claims
- No lessons learned
- Hypothetical outcomes only

Content Accessibility

Description

Evaluates how well technical concepts are explained for a broader audience
while maintaining scientific accuracy."

Scoring Logic

- Score 0: Technical concepts and content is too complex or oversimplified;
  Or this publication does not have technical concepts"
- Score 1: Technical concepts explained clearly at appropriate level

Positive Indicators

- Complex concepts broken down without loss of accuracy
- Technical terms properly introduced and explained
- Uses appropriate analogies or examples
- Builds understanding progressively
- Maintains scientific precision while being understandable

Negative Indicators

- Unexplained technical jargon
- Oversimplified to point of inaccuracy
- Requires PhD-level knowledge to understand
- Missing foundational explanations
- Poor balance between accuracy and accessibility

Technical Progression

Description

Assesses how well the content builds technical understanding progressively,
introducing concepts in a logical sequence that supports comprehension."

Scoring Logic

- Score 0: Poor progression of technical concepts
- Score 1: Logical progression that builds understanding effectively

Positive Indicators

- Concepts built up in logical sequence
- Prerequisites introduced before advanced topics
- Clear connection between concepts
- Appropriate pacing of technical content

- Background provided where needed

Negative Indicators

- Jumps to advanced concepts too quickly
- Missing important prerequisite concepts
- Unclear relationships between ideas
- Inconsistent technical depth
- Poor foundation building

Scientific Clarity

Description

Evaluates whether scientific accuracy is maintained while presenting content
in an accessible way."

Scoring Logic

- Score 0: Sacrifices scientific accuracy or is unclear
- Score 1: Maintains scientific accuracy while being clear

Positive Indicators

- Key scientific points accurately conveyed
- Important technical nuances preserved
- Complex ideas simplified without distortion
- Precise where it matters
- Clear distinction between simplification and actual mechanism

Negative Indicators

- Scientific inaccuracies in simplification
- Important technical details lost

- Misleading analogies or examples
- Over-generalization of specific findings
- Imprecise explanation of key concepts

Source Credibility

Description

Evaluates whether the blog properly references and cites its sources, clearly
identifies the origin of data/code/tools used, and provides sufficient version/
environment information for reproducibility. This helps readers validate claims,
trace information to original sources, and implement solutions reliably."

Scoring Logic

- Score 0: Missing or inadequate source citations, references, or version information
- Score 1: Clear, complete references and version information for all key claims, data, and tools"

Positive Indicators

- Links to original research papers or documentation for technical claims
- Clear identification of data sources and versions
- Specific versions of key tools/libraries stated
- Reproducibility information (environment, setup) provided
- Citations for statistics and quantitative claims

Negative Indicators

- Unsubstantiated technical claims

- Unclear origin of data or code
- Missing version information for critical tools
- Vague or incomplete references
- No reproducibility guidance when needed

Reader Next Steps

Description

Evaluates whether the publication provides clear guidance on what readers can do
after consuming the content. This includes suggested learning paths, topics to
explore, further reading materials, skills to practice, or actions to take. The
focus is on helping readers understand their potential next steps."

Scoring Logic

- Score 0: Missing or unclear guidance on what readers can do next
- Score 1: Clear suggestions for reader's next steps after this content

Positive Indicators

- Suggests specific next topics to learn
- Recommends practice exercises or projects
- Points to advanced concepts to explore
- Lists helpful resources for further learning
- Provides concrete action items for readers

Negative Indicators

- No guidance for reader's next steps
- Ends without suggestions for further learning
- Missing recommendations for practice

- No clear path for readers to progress
- Leaves readers uncertain about how to proceed

Uncommon Insights

Description

Evaluates whether the publication provides valuable insights that are either unique
(from personal experience/expertise) or uncommon (not easily found in standard
sources). Looks for expert analysis, real implementation experiences, or carefully
curated information that is valuable but not widely available."

Scoring Logic

- Score 0: Contains only common knowledge or easily available information
- Score 1: Provides unique or hard-to-find insights from experience or expertise

Positive Indicators

- Personal experiences from real implementations
- Novel combinations or analysis of existing concepts
- Expert interpretations of common challenges
- Unique perspectives on industry trends
- Specific insights from hands-on experience

Negative Indicators

- Generic explanations available in documentation
- Basic tutorials without unique insights
- Commonly repeated information

- Surface-level aggregation of existing content
- Standard explanations without added value

Technical Asset Access Links

Description

- Evaluates whether the publication provides links to access the technical asset
  (tool, dataset, model, etc.), such as repositories, registries, or download locations"
- Publication should include specific references to where readers can obtain or
  access the asset"

Scoring Logic

- Score 0: No mention of where or how to access the technical asset in the publication"
- Score 1: Clear references to where and how to access the technical asset included in
  the publication"

Positive Indicators

- Link to code repository (GitHub, GitLab, etc.) for software
- Reference to package registry (PyPI, npm, etc.) for libraries
- Dataset download links or access instructions
- Model hub links (HuggingFace, TensorFlow Hub, etc.)
- API endpoints or service access information
- Multiple access methods where applicable
- Version information included with references
- Data portal or catalog links for datasets
- Model card or model repository links
- Docker image location if containerized

Negative Indicators

- No access links or locations provided
- Vague mentions without specific URLs
- Missing version information
- No indication of how to access the asset
- Missing access credentials or request process for restricted assets

---

Installation and Usage Instructions

Description

Evaluates whether the publication provides clear instructions for installing and using
the tool, either directly in the publication or through explicit references to external
documentation. The key is that a reader should be able to quickly understand how to get
started with the tool."

Scoring Logic

- Score 0: No clear installation/usage instructions provided AND no explicit reference to
  external documentation containing these instructions. Or, external references are
  vague/broken/outdated."
- Score 1: Either contains clear installation/usage instructions directly in the publication
  OR provides specific, working references to external documentation (e.g., GitHub
  README, ReadTheDocs, PyPI page) that contains these instructions."

Positive Indicators

- Direct inclusion of installation commands (e.g., pip install, docker pull)
- Step-by-step usage instructions with code examples
- Specific links to installation guides (e.g., 'Installation instructions available at ')
- Direct references to package managers (e.g., 'Available on PyPI at ')
- Version-specific installation instructions
- Basic usage examples in publication with links to more comprehensive guides
- Clear prerequisites and dependency information
- Quick start guide with minimal working example
- Links to official documentation with specific sections for installation/usage

Negative Indicators

- No installation instructions or references to where they can be found
- Vague references like 'check our GitHub
- Missing crucial setup steps or prerequisites
- Installation/usage instructions without version information
- Incomplete instructions with no reference to complete documentation

---

Performance Characteristics and Requirements

Description

- Evaluates documentation of tool's performance characteristics
- Should cover resource requirements, scalability, and limitations
- Include benchmarks where applicable

Scoring Logic

- Score 0: Missing or unclear performance information
- Score 1: Clear documentation of performance characteristics and

requirements

Positive Indicators

- Resource requirements (CPU, memory, disk)
- Performance benchmarks
- Scalability information
- Known limitations and bottlenecks
- Hardware/software requirements

Negative Indicators

- No performance information
- Missing resource requirements
- No scalability guidance
- Unclear limitations

Maintenance and Support Status

Description

- Evaluates whether the publication clearly communicates the maintenance and support
  status of the technical asset (tool, dataset, model, etc.)"
- Should provide information about updates, versioning, and support channels"

Scoring Logic

- Score 0: No clear information about maintenance status or support channels"
- Score 1: Clear communication about maintenance status, versioning, and support options"

Positive Indicators

- Current version or release information
- Update/maintenance frequency stated
- Support channels or contact information provided
- Issue reporting process described
- Information about future updates/maintenance plans
- For datasets: data freshness, update schedule
- For models: retraining frequency, version compatibility

Negative Indicators

- No version or update information
- Unclear maintenance status
- Missing support information
- No indication of asset's current state

Access and Availability Status

Description

- Evaluates whether the publication clearly states how the technical asset can be
  accessed and used by others"
- Should explicitly state if it's open-source/public or has restricted access"

Scoring Logic

- Score 0: No clear information about access status and availability"
- Score 1: Explicitly states access status and provides relevant access information"

Positive Indicators

- Clear statement about public/private status
- For code: open-source status and repository access

- For datasets: data access methods and restrictions
- For models: model availability and distribution
- Contribution/collaboration guidelines if applicable
- Access request process for restricted assets

Negative Indicators

- Unclear availability status
- Missing access information
- Ambiguous terms about usage rights
- No information about how to obtain the asset

License and Usage Rights of the Technical Asset

Description

- Evaluates whether the publication clearly communicates the licensing terms and
  usage rights of the technical asset itself (not the publication). This includes
  software licenses for tools, data licenses for datasets, model licenses for AI
  models, etc."
- Should help readers understand what they are permitted to do with the asset
  once they obtain it"
- Note: This refers specifically to the license of the asset being described, not
  the license of the publication describing it"

Scoring Logic

- Score 0: No clear information about the technical asset's license or

usage rights provided"
- Score 1: Explicitly states the technical asset's licensing terms and usage rights"

Positive Indicators

- Clear statement of the asset's license type
- Asset's usage rights and restrictions clearly explained
- For code: software license (MIT, Apache, GPL, etc.) specified
- For datasets: data usage rights, sharing permissions, attribution requirements
- For models: model license, deployment restrictions, commercial usage terms
- Clear distinction between commercial vs non-commercial usage
- Asset's redistribution rights specified
- Asset's modification rights specified
- Reference to full license text for the asset

Negative Indicators

- No mention of the asset's license
- Unclear terms about what can be done with the asset
- Vague statements about asset usage permissions
- Missing important restrictions or requirements
- Only publication license mentioned, not asset license
- Ambiguous whether license refers to publication or asset

Contact Information of Asset Creators

Description

- Evaluates whether the publication provides information about how to contact

the creators/maintainers or the technical asset or get support, either directly
or through clear references to external channels"
- Should help readers know where to go for questions, issues, or support"

Scoring Logic

- Score 0: No contact or support information provided and no clear references to where
this information can be found"
- Score 1: Clear contact/support information provided in publication or specific
references to where this information can be found"

Positive Indicators

- Direct maintainer contact information
- Link to GitHub/GitLab issues
- Reference to discussion forums or chat channels
- Support email address
- Bug reporting guidelines
- Community support channels
- Office hours or support availability
- FAQ or troubleshooting resources

Negative Indicators

- No contact information provided
- No support channels mentioned
- No indication of where to get help
- Missing issue reporting process

Visual Tool Demonstration

Description

- Evaluates whether the publication includes visual aids to demonstrate the asset's
  usage and capabilities through demo videos, screenshots, diagrams, or step-by-step
  visual guides"
- Visual demonstrations help readers better understand the tool's interface,
  workflow, and expected outputs"

Scoring Logic

- Score 0: No visual demonstrations included or referenced in the publication"
- Score 1: Publication includes or references visual demonstrations through videos,
  screenshots, diagrams, or step-by-step visual guides"

Positive Indicators

- Demo video included or linked
- Screenshots of tool interface or outputs
- Step-by-step visual guides
- Diagrams showing tool workflow
- Visual examples of tool usage
- Before/after visuals showing tool effects
- Animated GIFs demonstrating functionality

Negative Indicators

- No visual aids included
- Text-only instructions without supporting visuals
- Missing visuals for key features or steps
- No demonstration of tool interface or outputs
