Supplementary File 5: AI-Based Implementation and Verification Protocols

Overview

This document provides complete protocols for AI-based implementation of the SPUR framework, including author self-assessment and peer reviewer verification procedures. Updated September 2025 to include field-specific baselines, conditional weighting, and enhanced replicability assessment.

Section 1: AI-Centric Implementation Model

Design Philosophy

The SPUR framework is designed for **AI-accessible implementation** rather than requiring human expert panels or institutional infrastructure. This approach enables:

- Universal accessibility for authors worldwide
- Rapid assessment turnaround (minutes vs. weeks)
- Consistent application of scoring criteria
- Cost-effective scalability
- Integration with existing publishing workflows

Implementation Stages

Stage 1: Author Self-Assessment
- Authors use AI (Claude, GPT, etc.) with standardized prompt
- Complete dimensional scoring with justifications
- Calculate SPUR score with conditional weighting logic
- Include assessment in manuscript submission

Stage 2: Peer Reviewer Verification
- Reviewers use verification prompt to validate author claims
- AI checks calculations, logic, and gaming indicators
- Reviewers provide human oversight on AI recommendations
- Accept, adjust, or reject author's SPUR score

Stage 3: Publication Integration
- Verified SPUR score included in abstract/metadata
- Field-normalized percentile ranking provided
- Dimensional breakdown available in supplementary materials
- Long-term citation tracking for framework validation

Section 2: Author Self-Assessment Protocol

When to Use

Authors should conduct SPUR self-assessment:
- Before manuscript submission (helps position novelty)
- During revision (demonstrates response to reviewer concerns)
- For preprint posting (signals uniqueness to readers)
- In grant applications (quantifies research innovation)

Required Information

Authors must provide to AI:
1. **Complete manuscript** (or detailed abstract for preliminary assessment)
2. **Research field/discipline** (primary and secondary)
3. **Target journal** (for field baseline context)
4. **Methodology description** (for innovation and replicability scoring)
5. **Related literature** (for positioning novelty claims)

Assessment Prompt

The complete author self-assessment prompt is provided in the main manuscript (Section: "Final Assessment and Call to Action"). It includes:

- Field-specific baseline identification
- Conditional weighting logic (Method Innovation ≥80 trigger)
- Enhanced replicability scoring (transparency vs. practical barriers)
- Step-by-step calculation with conditional adjustments
- Gaming resistance checks

Location: Main manuscript, after "Final Assessment and Call to Action" section, marked with copy delimiters.

Section 3: Peer Reviewer Verification Protocol (UPDATED)

Purpose

Peer reviewers verify author-claimed SPUR scores to:

- Ensure accurate dimensional scoring
- Detect gaming or score inflation
- Validate conditional weighting application
- Confirm field baseline appropriateness
- Check replicability barrier assessment
- Provide independent quality control

Verification Prompt for Peer Reviewers

COPY EVERYTHING BELOW THIS LINE FOR PEER REVIEW VERIFICATION

SPUR Score Verification Protocol for Peer Reviewers

You are verifying an author's self-assessed SPUR (Scientific Paper Uniqueness Ranking) score for a submitted manuscript. Your role is to validate the claimed score, check for gaming, and ensure proper application of the framework's three critical enhancements.

Manuscript Information Needed

1. **Author's claimed SPUR score**: [Insert score]
2. **Author's dimensional breakdown**: [Insert 7 scores]
3. **Research field**: [Insert primary/secondary discipline]
4. **Author's field baseline used**: [Insert baseline mean]
5. **Manuscript title and abstract**: [Paste here]
6. **Author's SPUR justification**: [Paste their explanation]

Verification Steps

STEP 1: Field-Specific Baseline Verification

Check if author used appropriate baseline for their discipline:

Expected baselines (mean SPUR scores for "typical" papers):
- Biomedical Sciences: 45-50/100
- Physical Sciences: 42-48/100
- Social Sciences: 48-52/100
- Mathematics: 40-45/100

- Humanities: 38-45/100

Verification questions:
1. Did the author correctly identify their primary research field?
2. Is the baseline they used within the expected range?
3. For interdisciplinary work, did they justify baseline selection?
4. If they used field normalization (z-score), was the calculation correct?

Red flags:
- Using baseline from wrong discipline to inflate percentile ranking
- Cherry-picking lowest baseline to appear more unique
- Ignoring field context entirely

STEP 2: Conditional Weighting Verification

Check if conditional weighting was properly applied:

Rule: IF Methodological Innovation score ≥ 80/100, THEN:
- Methodological Innovation weight: 20% → 25%
- Replicability & Transparency weight: 10% → 5%

Rationale: Highly innovative methods are inherently harder to replicate, so the framework automatically reduces replicability penalties for breakthrough methodologies.

Verification questions:
1. What is the author's Methodological Innovation score? [Check]
2. If ≥80, did they apply 25% weight to Method Innovation? [Yes/No]
3. If ≥80, did they apply 5% weight to Replicability? [Yes/No]
4. If <80, did they use standard weights (20% and 10%)? [Yes/No]
5. Is the conditional trigger justified by truly innovative methodology?

Red flags:
- Inflating Method Innovation score to 80+ just to trigger weighting shift
- Applying conditional weights when Method Innovation <80
- Not applying conditional weights when Method Innovation legitimately ≥80

STEP 3: Replicability Barrier Assessment

Check if replicability scoring distinguishes transparency from practical barriers:

The framework now recognizes two types of replication challenges:

Type A: Transparency Issues (penalized normally)
- Missing data
- Inadequate method descriptions
- Proprietary tools without alternatives
- Selective reporting
- **These should lower replicability scores**

Type B: Legitimate Barriers (not penalized when Method Innovation ≥80)
- Novel equipment not yet widely available
- Cutting-edge techniques requiring specialized training
- Unique observational opportunities (eclipses, rare events)
- Computational requirements beyond standard resources
- **These trigger conditional weighting protection**

Verification questions:
1. Did the author distinguish between transparency and legitimate barriers?
2. If claiming high Method Innovation, are replication barriers justified?
3. Are there any transparency issues being hidden behind "innovation" claims?
4. Is the replicability score reasonable given the methodology?

Red flags:
- Claiming "innovative so can't replicate" to hide poor documentation
- Conflating legitimate barriers with sloppy methodology
- Over-penalizing genuinely innovative work for inherent barriers

---

STEP 4: Calculate Independent SPUR Score

Using the author's manuscript, score each dimension independently:

1. Methodological Innovation (0-100): [Your score]
- Novel techniques, tools, or approaches
- Advancement beyond current methods
- Transferability to other research

2. Conceptual Originality (0-100): [Your score]
- New theoretical frameworks
- Unexpected connections
- Paradigm-challenging insights

3. Empirical Scope & Scale (0-100): [Your score]
- Data comprehensiveness
- Temporal/spatial coverage
- Population/sample representativeness

4. Societal Impact Potential (0-100): [Your score]
- Policy relevance
- Economic implications
- Social/environmental benefit
- Stakeholder reach

5. Cross-Disciplinary Integration (0-100): [Your score]
- Methods from multiple fields
- Synthetic theoretical frameworks
- Bridges between disciplines

6. Replicability & Transparency (0-100): [Your score]
- **Assess transparency issues** (data, code, methods)
- **Note legitimate barriers separately**
- Consider whether conditional weighting applies

7. Theoretical Advancement (0-100): [Your score]
- New predictions or explanations
- Generalizable principles
- Theoretical coherence

STEP 5: Apply SPUR Formula with Conditional Logic

Calculate base score with proper weights:

```
IF Methodological_Innovation ≥ 80:
   Method_Weight = 0.25
   Replicability_Weight = 0.05
ELSE:
   Method_Weight = 0.20
   Replicability_Weight = 0.10

Base_SPUR = (Methodological × Method_Weight) +
        (Conceptual × 0.18) +
```

```
        (Empirical × 0.15) +
        (Societal × 0.15) +
        (CrossDisciplinary × 0.12) +
        (Replicability × Replicability_Weight) +
        (Theoretical × 0.10)
```

Apply impact multiplier:

```
Multiplier = 1 + (0.3 × Societal_Impact / 100)
Final_SPUR = Base_SPUR × Multiplier
```

Your calculated SPUR score: [Insert]

STEP 6: Gaming Detection Checks

Examine the manuscript and author's justification for these gaming patterns:

1. Vocabulary Manipulation Gaming
- Using jargon complexity to appear novel
- **Check**: Is complexity justified by actual innovation?

2. Method Combination Gaming
- Combining standard methods and claiming innovation
- **Check**: Is synthesis genuinely novel or just procedural?

3. False Interdisciplinary Claims
- Superficial citations across fields
- **Check**: Are methods truly integrated or just referenced?

4. Impact Exaggeration
- Overstating policy/societal relevance
- **Check**: Is impact plausible and well-justified?

5. Complexity Obfuscation
- Obscuring simple methods to appear sophisticated
- **Check**: Could methods be described more simply?

6. Baseline Shopping

- Selecting field baseline to maximize percentile
- **Check**: Is discipline classification accurate?

7. Conditional Weighting Gaming
- Inflating Method Innovation to ≥80 just for weight shift
- **Check**: Is ≥80 score genuinely warranted?

STEP 7: Comparison and Recommendation

Score Comparison:
- Author's SPUR: [X]
- Your verified SPUR: [Y]
- Difference: [Y - X]

Acceptable differences: ±5 points (subjective interpretation range)

Recommendation:

ACCEPT if:
- Difference ≤5 points
- No gaming indicators detected
- Field baseline appropriate
- Conditional weighting correctly applied
- Replicability barriers properly assessed

ADJUST if:
- Difference 6-10 points
- Minor gaming indicators (vocabulary, impact slight exaggeration)
- Baseline defensible but suboptimal
- Conditional logic applied but marginal (Method Innovation 78-82 range)
- Suggest revised score with justification

REJECT if:
- Difference >10 points
- Clear gaming detected (false interdisciplinary, baseline shopping)
- Wrong baseline intentionally used
- Conditional weighting misapplied
- Transparency issues hidden behind innovation claims
- Require author to reassess with your feedback

STEP 8: Provide Feedback to Author

Your verification report should include:

1. **Verified SPUR score** (accept author's, or provide adjusted score)
2. **Dimensional score adjustments** (which dimensions changed and why)
3. **Field baseline confirmation** (correct or needs revision)
4. **Conditional weighting check** (properly applied or not)
5. **Replicability assessment** (transparency vs. barriers clarification)
6. **Gaming concerns** (if any detected)
7. **Justification** (2-3 sentences explaining verification outcome)

Example feedback format:

> "The author's claimed SPUR score of 87.3 has been verified with minor adjustment to 84.6. The Methodological Innovation score of 82 appropriately triggers conditional weighting (25%/5%), and the replication barriers are legitimate given the novel neuroimaging technique. However, the Societal Impact score was reduced from 88 to 78 as the policy applications, while promising, are not yet empirically demonstrated. The field baseline for Biomedical Sciences (47.5) is appropriate. No gaming indicators detected. **Recommendation: ACCEPT with adjusted score of 84.6/100 (93rd percentile).**"

END OF PEER REVIEWER VERIFICATION PROTOCOL

Section 4: Quality Assurance for AI Assessments

AI Model Requirements

Recommended AI models for SPUR assessment:
- Claude 3.5+ (Anthropic)
- GPT-4+ (OpenAI)
- Gemini Ultra (Google)

Minimum capabilities needed:
- Long context window (50k+ tokens for full manuscripts)
- Numerical reasoning accuracy
- Conditional logic execution
- Scientific literature comprehension

- Gaming pattern recognition

Human Oversight Checkpoints

While AI conducts primary assessment, humans should verify:

1. **Author stage**: Review AI's dimensional scores for face validity
2. **Peer review stage**: Independently judge 1-2 key dimensions
3. **Editorial stage**: Spot-check calculations and percentile rankings
4. **Post-publication**: Track citations to validate predictive power

Known AI Limitations

- May over-credit vocabulary complexity as innovation
- May under-appreciate subtle theoretical advances
- Field baseline selection requires human domain knowledge
- Societal impact projection inherently speculative
- Cannot access paywalled comparison literature

Mitigation: Authors and reviewers provide context AI cannot access.

Section 5: Documentation Standards

Required Author Documentation

When submitting SPUR self-assessment, authors must include:

1. **AI model used** (name and version)
2. **Complete prompt** (copy of assessment prompt used)
3. **Full AI response** (dimensional scores with justifications)
4. **Author's verification** (statement agreeing or adjusting AI scores)
5. **Field baseline source** (citation or empirical basis)
6. **Conditional weighting notation** (if triggered)
7. **Replicability barriers** (transparency issues vs. legitimate barriers)

Transparency Requirements

- Authors must disclose AI use in SPUR assessment
- Full calculation audit trail in supplementary materials
- Dimensional score justifications (2-3 sentences each)
- Any deviations from standard procedure explained

Section 6: Integration with Academic Publishing

Submission Integration

In manuscript abstract:
> "SPUR Score: 86.2/100 (94th percentile, Biomedical Sciences)"

In methods/study design:
> "Research uniqueness assessed using SPUR framework (Miller, 2025), with dimensional scores provided in Supplementary Materials."

In supplementary materials:
- Complete SPUR calculation worksheet
- AI assessment transcript
- Peer reviewer verification (post-acceptance)

Journal Adoption Pathway

Phase 1: Voluntary adoption (2025-2026)
- Authors optionally include SPUR scores
- Journals acknowledge but don't require
- Build validation dataset

Phase 2: Encouraged reporting (2027-2028)
- Journals request SPUR in author guidelines
- Peer reviewers asked to verify
- Databases begin tracking scores

Phase 3: Standard practice (2029+)
- SPUR required for novelty claims
- Integrated into impact metrics
- Funding agencies reference scores

Section 7: Ethical Considerations

Avoiding Misuse

SPUR should NOT be used to:

- Reject papers solely based on low scores
- Compare researchers for hiring/promotion
- Rank journals or institutions (see I-SPUR for institutional assessment)
- Replace nuanced peer review judgments
- Penalize replication studies (which legitimately score lower)

Appropriate Uses

SPUR is designed to:
- Help authors position novelty claims objectively
- Give reviewers structured assessment tool
- Identify highly unique research for recognition
- Track innovation trends over time
- Inform meta-research on scientific progress

Gaming Mitigation

The framework includes built-in gaming resistance:
- Multi-dimensional prevents single-axis inflation
- Impact multiplier caps at 1.3× (not 2× or 5×)
- Conditional weighting only for genuinely high innovation
- Replicability distinguishes transparency from barriers
- Peer verification catches score inflation

Section 8: Version Control and Updates

Current Version: 1.1 (September 2025)

Changes from v1.0:
- Added field-specific baselines (5 discipline categories)
- Implemented conditional weighting (Method ≥80 trigger)
- Enhanced replicability scoring (transparency vs. barriers)
- Updated peer reviewer verification protocol
- Added gaming resistance checks for conditional weighting

Future updates planned:
- Expand field baselines to 15+ disciplines (empirical calibration)
- Refine conditional weighting threshold (currently 80, may adjust to 75 or 85)
- Add temporal decay factors for citation validation
- Develop institutional extension (I-SPUR)

Feedback welcome: rrobbyymiller@gmail.com

Conclusion

The SPUR framework's AI-accessible implementation enables universal, consistent, and transparent research uniqueness assessment. The three critical enhancements (field baselines, conditional weighting, replicability barriers) are now fully integrated into both author self-assessment and peer reviewer verification protocols.

By combining AI efficiency with human oversight, SPUR provides a practical solution to objectively evaluating research novelty at scale while maintaining gaming resistance and cross-field comparability.

Document prepared by: Robert Miller, Independent Researcher, Sydney, Australia
Date: October 1, 2025
Contact: rrobbyymiller@gmail.com
ORCID: 0009-0006-4120-313X