

## REVIEWER'S REPORT

Manuscript No.: **IJAR-52461**

Date: **24/06/2025**

**Title: ImageStory: Enhanced Cognitive Visual Narrative**

### Recommendation:

Accept as it is .....

**Accept after minor revision (✓)**

Accept after major revision .....

Do not accept (*Reasons below*) .....

| Rating         | Excel. | Good | Fair | Poor |
|----------------|--------|------|------|------|
| Originality    |        | ✓    |      |      |
| Techn. Quality |        | ✓    |      |      |
| Clarity        |        |      | ✓    |      |
| Significance   |        | ✓    |      |      |

Reviewer Name:

**Yuniana Cahyaningrum, S.Kom., M.Kom.**

Date:

**24/06/2025**

### Reviewer's Comment for Publication.

This manuscript presents the Enhanced Cognitive Visual Narrative System (ECVNS), a state-of-the-art multimodal artificial intelligence framework designed for automated visual storytelling. The system integrates several state-of-the-art deep learning models including OWLv2 for object detection, BLIP for image captioning and visual question and answering, CLIP for emotional analysis, and ViLT for scene understanding. This work contributes to the emerging field of multimodal AI and has applications in content creation, accessibility, education, and entertainment. With minor improvements (especially clarification of sample numbers, statistical reporting, and language refinement), this work could make a valuable contribution to the IJAR readership. Therefore, I recommend acceptance after minor revisions.

## Detailed Reviewer's Report

### Strengths

1. Relevant topic  
Investigating how to combine computer vision techniques with natural language generation to create an integrated system that can understand visual content at multiple semantic levels and translate this understanding into creative storytelling.
2. Clear research aim  
The guiding questions and hypotheses are stated at the beginning, so that the reader can follow how the next steps will be.
3. Sound instruments  
This system achieves high accuracy in object detection, scene understanding, and emotional inference, producing narratives that demonstrate technical precision and creative quality..
4. Practical implications

# International Journal of Advanced Research

**Publisher's Name: Jana Publication and Research LLP**

*www.journalijar.com*

---

## **REVIEWER'S REPORT**

Recommendations for governments, educational boards, and international organizations to adopt and implement this model within the next five years.

5. Ethical transparency

This paper contributes to the growing field of multimodal AI and has applications in content creation, accessibility, education, and entertainment.

### **Weaknesses**

1. Incomplete statistics

Some data from manuscript no calculate..

2. Table and Graph

There is no graph to inform consent the data tabulation.

3. Language polish

Minor grammar slips distract from the argument; a quick copy-edit would fix this.

4. Reference consistency

A few URLs are incomplete and year formats vary. Aligning all entries with APA 7 will enhance professionalism. The reference use must be up to date (Last 5 years).