

# Large Language Models for Detecting Gambling Advertisement Images to Enhance the Efficiency of the Creative Review Process

**Presented by:**

Jayesh Santosh Asawa

Work by: Edward L Martis, Jayesh Santosh Asawa

# Challenges in Creative Review

- Multiple Advertising Policies (e.g.: No Gambling, No Adult, No Crypto etc).
- Current process relies on third-party APIs (e.g. Cloud Vision) for initial image filtering.
  - These APIs struggle with specialized policies - e.g., 70% of gambling images marked safe.
- To uphold the brand integrity and maintain trust and ethical standards, manual reviews are most reliable.
  - The process slow, costly, and error-prone.

# Gambling Category

- **Explored CNN-based Transfer Learning (VGG-16) and Vision Transformer (ViT):**
  - Fine-tuned pre-trained models to classify gambling vs. non-gambling images.
- **Experimented with LLM-based approaches:**
  - Used LLAMA-Vision-11B with prompt engineering and LLM2Vec encoder for zero-shot and contextual classification improvements.
- **Ensemble Model:**
  - Combined outputs from CNN, ViT, and LLM2Vec using logistic regression for best performance.
- **Comparative Analysis:**
  - Evaluated all models on FNR, FPR, and F1-score.

# Dataset

- **Source:** Proprietary data from Samsung's SSP ad inventory logs. Covers casino games, real-money games, cryptocurrency gambling, lottery, and sweepstakes ads.
- **Initial Size:** 50K unique images, refined to 30K labeled images.
- **Labeling:** Binary classes: Gambling (4%) vs Non-Gambling (96%).
- **Preprocessing:** Standardized sizes (64x64, 128x128, 224x224).
- **Split:** Train (60%), Validation (20%), Test (20%).

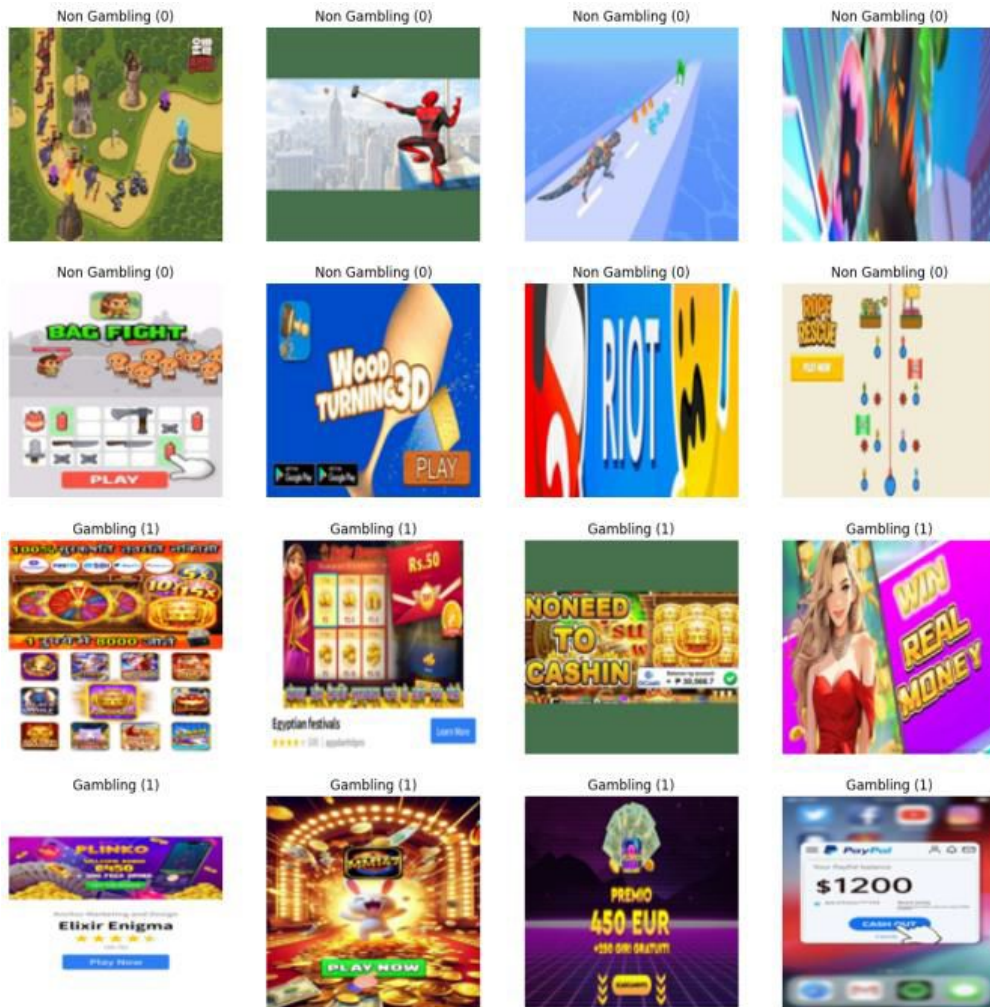


Fig. 1. Sample Creative Ads images in Dataset with Labels of Gambling and Non Gambling.

# VGG-16 (CNN with Transfer Learning)

**Architecture:** 13 convolutional + 3 fully connected layers (16 total), uses 3x3 filters for fine-grained details.

**Pre-training:** Trained on ImageNet, leveraged for gambling image classification via **transfer learning**.

## Tuning:

- Replaced final layer with 2-class output (Gambling / Non-Gambling).
- Used **Adam optimizer**, small learning rate (0.0001).
- **Class Imbalance Handling:** Applied **class weights (1:40)** for gambling vs non-gambling.

## Training:

- Started with frozen convolutional layers, experimented with partial fine-tuning.
- Input sizes: 64x64, 128x128, 224x224 tested.
- Validation accuracy plateaued after ~40 epochs.
- Evaluated using **FNR, FPR, and F1-score**.

# Vision Transformer (ViT)

## Architecture:

- Treats image as a sequence of patches of size 16x16, uses self-attention for global context.
  - Patches, Linear embedding + positional encoding, Transformer encoder.
- Captures global image context through self-attention, unlike CNNs that focus on localized features.

## Adaptation:

- Fine-tuned pre-trained ViT model on gambling dataset.
- Trained for 100 epochs; validation accuracy plateaued after ~20 epochs.

## Performance:

- Metrics (Precision, Recall, F1-score) comparable to VGG-16.
- Observed overfitting beyond 20 epochs.

**Advantage:** Handles varied image contexts effectively using global attention.

# Using LLMs - Approach 1

## Process:

- Used LLAMA Vision 11B model, got some prompt engineering ideas with the help of Deepseek-R1 to identify gambling related items in the image.
- Input: Image for classification + Prompt
- Model outputs structured JSON (Gambling = 1, Not Gambling = 0).

## Results:

- **False Positives:** 19.21%
- **False Negatives:** 8.47%

## Observation:

- LLM performed **poorly for pure discriminative / classification tasks** for our use case compared to CNN and ViT.



# LLM Approach 2

- **Steps:**

- Used **Manual Review + DeepSeek-R1** to create prompts.
- Extracted image descriptions using Vision LLM, fed into **LLM2Vec encoder** for contextual embeddings.
- Trained **Logistic Regression** (SAGA solver, balanced class weights) on encoded vectors.

- **Advantages:**

- Captures textual patterns describing gambling cues.
- **Significant performance improvement** over Approach 1 and baseline models (VGG16, ViT).

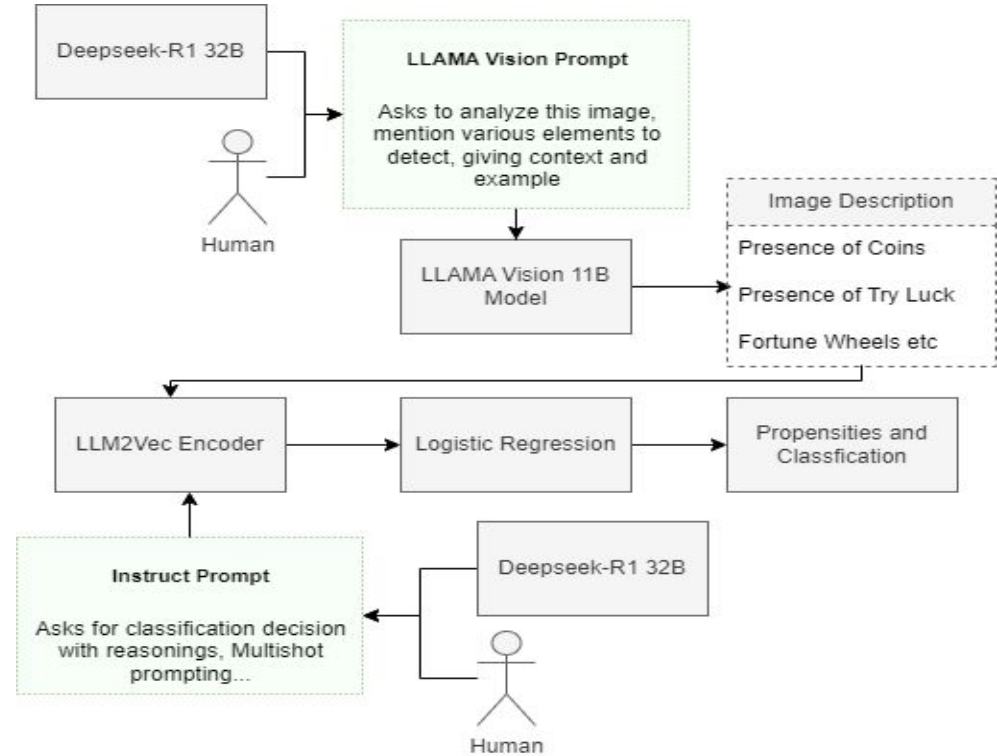


Fig. 2. Approach 2 For LLM2Vec based classification

### LLAMA Vision 11B Prompt:

Analyze this image and list all observable elements that could suggest it is related to risk-for-reward mechanics, monetary wagers, or casino-style activities. If absent, list elements that suggest generic advertising, entertainment, or other purposes.

Focus only on these categories: Monetary/Financial Elements: Casino chips, coins, banknotes, piles of cash Currency symbols (\$, €, ¥), free spins" or "bonus" text. Prize amounts, jackpot displays, "win" or "wager" in text

Game Mechanics: Playing cards, dice, roulette wheels, slot machines. Poker tables, numbered balls (e.g., lottery), sports odds. Progress bars, "level up" indicators, timer countdowns

Ad/Creative Elements (if no gambling cues): Brand logos, app store badges, celebrity endorsements Neutral graphics (e.g., landscapes, abstract art) Call-to-action text ("Download Now," "Try Free")

Rules:

Only list items verbatim (e.g., "red poker chips," "\$100 bill," "slot machine animation"). No summaries (avoid phrases like "this is gambling-related"). No assumptions—only include visible/textual elements.

Example Output:

Three golden casino chips stacked

Text: "Daily Bonus: 500 Coins" Green felt table with card symbols (:spades::hearts::diamonds::clubs:) Spin-the-wheel animation in background

Fig. 3. LLAMA Vision 11B Prompt

#### Instruction Prompt:

Classify as gambling or non-gambling. Some examples for gambling images:

Winning & Jackpot Focused:

"A euphoric gambler celebrating a massive slot machine jackpot win, coins pouring out, with a bright 'WINNER' sign flashing

and a disclaimer: 'Gambling involves risk. Play responsibly.'"

"A poker player revealing a royal flush, opponents groaning, with a pile of cash and chips pushed toward them—text overlay: 'Know when to walk away.'"

"A lottery ticket being scratched off, revealing a '\$1,000,000' prize, with fine print at the bottom: 'Odds of winning vary. Bet with caution.'"

Risk & Luck-Based Imagery:

"A roulette wheel spinning, a nervous bettor watching the ball, with a casino sign in the background: 'Luck is a factor. Don't chase losses.'"

"A gambler hesitating before placing a high-stakes bet, a shadowy casino backdrop, with text: 'Risk is part of the game. Set limits.'"

"A slot machine screen displaying 'BONUS ROUND—YOU WON \$500!' with a small disclaimer: 'Results are random. Play for fun, not profit.'"

"You've Already Won" or "Try Your Luck" Themes:

"A casino billboard glowing with 'TRY YOUR LUCK TODAY!' alongside images of dice, cards, and a tiny 'Terms and conditions apply.'"

"A smartphone screen showing a 'CONGRATS! YOU WON \$200 CASH!' pop-up from a gambling app, with a 'Claim Prize' button and risk warning in small text."

"A gambler holding a golden ticket with 'INSTANT CASH PRIZE' stamped on it, surrounded by slot machines, and a subtle 'Gambling can be addictive.'"

"A blackjack table with a dealer smiling, pushing chips toward a player, text overlay: 'Feeling lucky? Remember—the house always has an edge.'"

Fig. 4. Instruction Prompt for LLM2Vec



Fig. 5. Sample Gambling Image

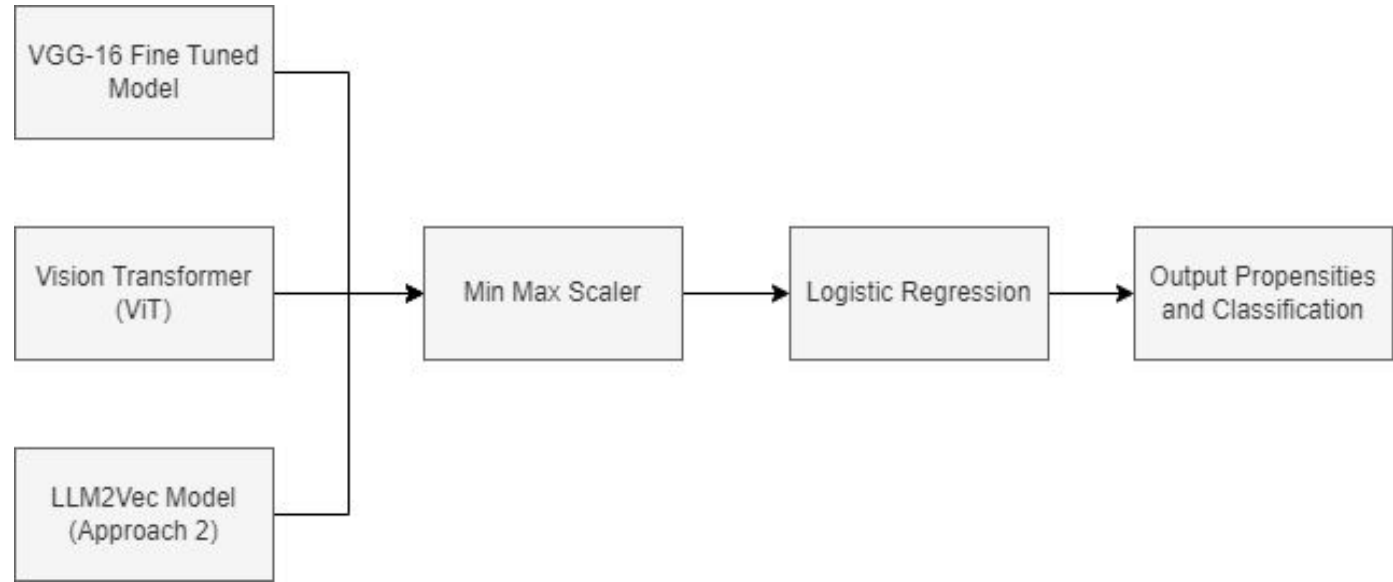
Response:

Based on the presence of the words ""Rewards"" and the wheel of fortune, the image appears to be advertising a gambling game for older aged people. The presence of gold coins at the bottom of the image adds to the suspicion that this is a gambling game for older players. The image is likely designed to entice mature players by describing the rewards they will receive after accumulating certain amounts of coins. The presence of a balloon and a gift box further reinforces this idea.

# Ensemble Model

We combined the output propensities got from the VGG16, ViT and LLM2Vec model outputs and applied min-max normalization over the propensities and then trained a logistic regression classifier on them. This technique further improved the results.

Fig. 6. Ensemble Model



# Results

Model	Dataset	FPR (%)	FNR (%)	F1-Score
<b>VGG16</b>	Test	2.32	12.16	0.8208
	Val	2.06	12.31	0.8388
<b>ViT</b>	Test	3.22	12.72	0.7342
	Val	3.55	12.56	0.7148
<b>LLM2Vec</b>	Test	1.98	11.68	0.8516
(Approach 2)	Val	1.83	11.91	0.8414
<b>Ensemble</b>	Test	1.59	11.02	0.8806
	Val	1.57	10.97	0.8923

Table 1: Results from different models

# Conclusion

- The manual review team need not spend time on 91% of images predicted as “Not gambling” and 4% of the images predicted as “Gambling”.
- Our model, therefore, saves 95% of the time and effort with only 5.36% of gambling images (i.e. 17 gambling misclassified out of (17+59+241=317)) and 0.61% of non-gambling images misclassified.

Predicted Group	Actual 0	Actual 1	% Miss	% of Total
Not Gambling	6,061	17	0.61%	91%
Manual Check	234	59	-	4%
Gambling	39	241	5.36%	4%

Table 2: Final Results of Validation Set with Manual Review

# Thank You

You can reach out to us at:

- Jayesh Santosh Asawa ([jayesh.asawa@samsung.com](mailto:jayesh.asawa@samsung.com), [jayeshasawa1@gmail.com](mailto:jayeshasawa1@gmail.com))
- Edward L Martis ([e.martis@samsung.com](mailto:e.martis@samsung.com))