



# WhatsApp Bot

## Intelligent Group Moderation Solution

Automated content moderation, user verification, and analytics for WhatsApp groups - powered by AI

IHL Presentation | 2026

## Section 01

# Agenda

This presentation provides a comprehensive overview of the WhatsApp Bot solution, covering its architecture, capabilities, and deployment options.

1

## System Architecture

Technical overview of components and technologies

2

## Bot Functionalities

Complete feature set and capabilities

3

## Moderation Pipeline

5-layer content analysis system

4

## Sensitive Topics

8 monitored content categories

5

## Deployment Options

Local, cloud, and mobile setups

## Key Highlights

- **Multi-layer AI-powered content moderation** - Real-time analysis using OpenAI
- **CAPTCHA-based user verification** - Prevents bots and spam accounts
- **Real-time web dashboard** - Monitor status, logs, and analytics
- **Weekly AI-generated reports** - Trending topics and engagement metrics
- **Flexible deployment** - Run on local machine, cloud, or smartphone

## Section 02

# System Architecture

## Technology Stack

Node.js 18+

Baileys (WhatsApp Web)

Express.js

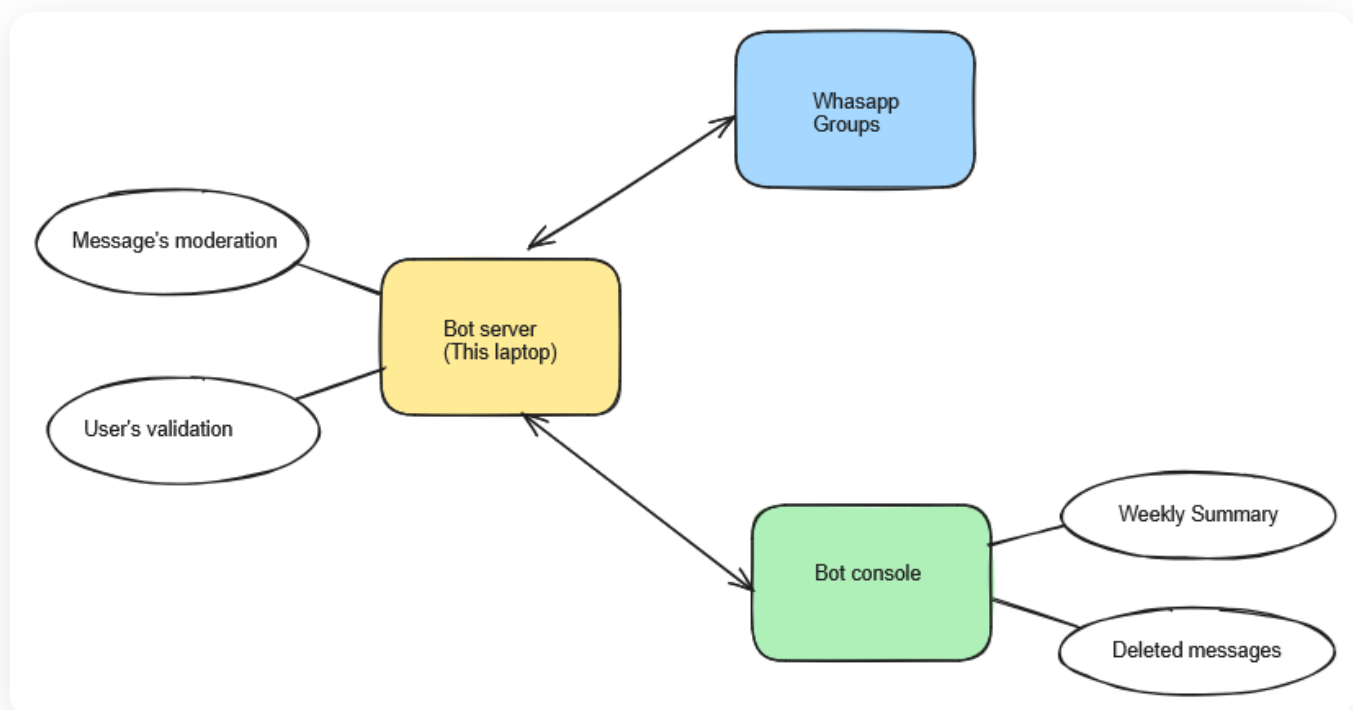
SQLite3

OpenAI GPT-4o-mini

Vision API

Moderation API

## Architecture Diagram



## Section 03

# Bot Functionalities



## Intelligent Content Moderation

Multi-layer AI system that detects spam, toxic content, and sensitive topics in real-time. Automatically removes violations and notifies users.



## CAPTCHA Verification

New members must solve an image-based CAPTCHA to stay in the group. Prevents bots and spam accounts. Multilingual support (EN/NL).



## Web Dashboard

Real-time monitoring interface showing connection status, moderation logs, and group analytics. Accessible via browser at port 3000.



## Weekly Analytics

AI-generated reports with trending topics, engagement metrics, most active days, and conversation summaries for each monitored group.



## Message Restoration

Administrators can review and restore accidentally deleted messages through the dashboard. Full audit trail maintained for all actions.



## Detailed Logging

Complete audit trail of all moderation actions including timestamps, violation types, and original message content for compliance review.

## Section 04

## 5-Layer Moderation Pipeline

The bot uses a multi-layer moderation pipeline that processes messages in sequence, from fast local checks to advanced AI analysis.

1

### Spam Detection

Checks for excessive links (>3), caps ratio (>70%), and repeated messages within 60 seconds

Local Rules

2

### Keyword Detection

Fast regex scan for 28 sensitive topic keywords across 8 categories

Regex Scan

3

### Toxic Content Analysis

Detects hate speech, harassment, violence, self-harm, and explicit content

OpenAI Moderation

4

### Sensitive Topic Intelligence

GPT-4o-mini analyzes nuanced discussions with confidence scoring (>0.7 threshold)

GPT-4o-mini

5

### Image Content Analysis

Vision API scans images for policy violations with 0.6 confidence threshold

Vision API

## Moderation Actions

- **Message Deletion:** Violating content is removed from the group immediately
- **User Notification:** Private message sent explaining the violation reason
- **Audit Logging:** Full details stored in database for review and restoration
- **Admin Exemption:** Group administrators bypass all moderation checks

## Section 05

## 8 Sensitive Topics Monitored

The bot monitors and moderates discussions on these sensitive topic categories to maintain a respectful group environment:



### Politics

Politicians,  
parties, elections,  
voting,  
government  
discussions



### Religion

God, church,  
mosque, temple,  
bible, quran,  
conversion topics



### Caste

Caste system,  
social divisions,  
reservation  
debates



### Gender

Feminism  
debates,  
misogyny,  
patriarchy  
arguments



### LGBTQ+

Sexual  
orientation, same-  
sex marriage  
debates



### Racism

Skin color,  
regionalism,  
ethnic  
stereotypes



### Health Misinfo

Anti-vax content,  
conspiracy  
theories, COVID  
misinformation



### Abortion

Pro-life vs pro-  
choice debates  
and related  
content

### Detection Methods




- **Keyword Matching:** Fast regex scan for known trigger words and phrases
- **AI Context Analysis:** GPT-4o-mini understands nuanced discussions beyond simple keywords

- **Image Recognition:** Vision API detects sensitive content in shared images
- **Confidence Thresholds:** Configurable sensitivity levels to reduce false positives

## Section 06

## Deployment Options & Costs

The WhatsApp Bot can be deployed in three main configurations, each with different cost and operational characteristics.

 Local Machine	 Cloud Server	 Smartphone
Monthly Cost <b>\$6-20</b> Electricity + OpenAI	Monthly Cost <b>\$6-25</b> VPS + OpenAI	Monthly Cost <b>\$2-6</b> OpenAI only
<b>✓ Advantages</b> <ul style="list-style-type: none"><li>• No hosting fees</li><li>• Full data control</li><li>• Easy initial setup</li><li>• Direct log access</li></ul> <b>✗ Disadvantages</b> <ul style="list-style-type: none"><li>• PC must run 24/7</li><li>• Electricity costs</li><li>• Manual restarts needed</li><li>• Internet dependency</li></ul>	<b>✓ Advantages</b> <ul style="list-style-type: none"><li>• 24/7 guaranteed uptime</li><li>• Auto-restart on failure</li><li>• Remote access anywhere</li><li>• Scalable resources</li></ul> <b>✗ Disadvantages</b> <ul style="list-style-type: none"><li>• Monthly hosting fees</li><li>• Technical setup required</li><li>• Data on third-party</li><li>• Server knowledge needed</li></ul>	<b>✓ Advantages</b> <ul style="list-style-type: none"><li>• Always connected</li><li>• Low power usage</li><li>• Portable solution</li><li>• No extra hardware</li></ul> <b>✗ Disadvantages</b> <ul style="list-style-type: none"><li>• Limited processing</li><li>• Complex setup (Termux)</li><li>• Battery drain</li><li>• Storage constraints</li></ul>

### Recommended Setup by Use Case

Use Case	Recommended	Reason
Personal / Small Community	<b>Local Machine</b>	Simple setup, full control, minimal costs
Business / Professional	<b>Cloud Server</b>	Reliability, multiple groups, remote management
Testing / Development	<b>Local Machine</b>	Ideal for initial configuration before production
Remote / Mobile Setup	<b>Smartphone</b>	Portability for tech-savvy users with limited resources

Appendix A

# OpenAI API Cost Analysis

## Client Parameters

Parameter	Value
WhatsApp Groups	15 groups
Text Messages per Day per Group	40 messages
Average Message Size	120 characters
Images per Day per Group	6 images

## Monthly Volume

18,000	2,700
Text Messages/Month	Images/Month

## Moderation Pipeline & API Usage

Step	Function	API	Cost
1	Spam Detection	Local rules	Free
2	Keyword Detection	Local regex	Free
3	Toxic Content Check	Moderation API	Free

Step	Function	API	Cost
4	Sensitive Topic Analysis	GPT-4o-mini	Per token
5	Image Analysis	GPT-4o-mini Vision	Per token

## Appendix A (Continued)

## Monthly Cost Calculation

### Token Consumption Estimate

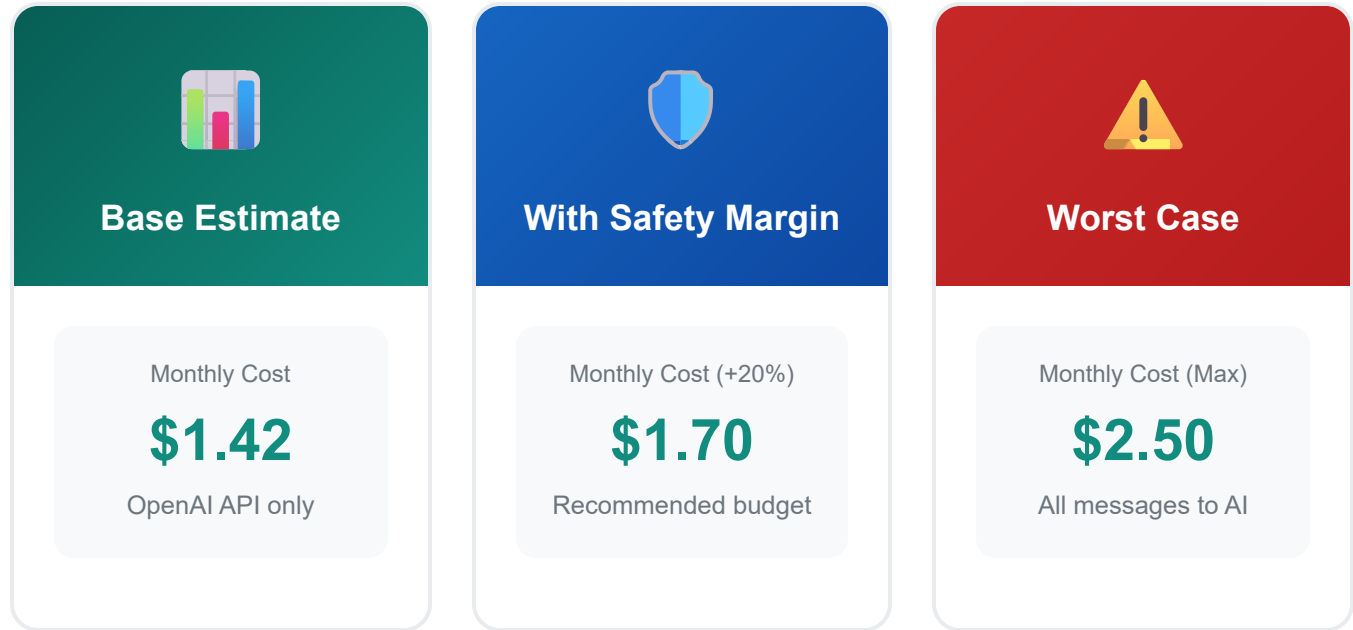
Assuming ~90% of messages pass local filters and have >20 characters:

API Call Type	Calls/Month	Input Tokens	Output Tokens
Moderation API	18,000	-	-
GPT-4o-mini (text)	~16,200	4,050,000	810,000
GPT-4o-mini Vision	2,700	1,039,500	270,000
<b>TOTAL</b>	<b>36,900</b>	<b>5,089,500</b>	<b>1,080,000</b>

### Cost Breakdown (GPT-4o-mini Pricing)

Component	Tokens	Rate	Cost
Text Analysis - Input	4,050,000	\$0.15/1M	\$0.61
Text Analysis - Output	810,000	\$0.60/1M	\$0.49
Image Analysis - Input	1,039,500	\$0.15/1M	\$0.16
Image Analysis - Output	270,000	\$0.60/1M	\$0.16
Moderation API	18,000 calls	Free	\$0.00
<b>TOTAL MONTHLY</b>			<b>\$1.42</b>

## Cost Summary



### Notes

- **Moderation API is FREE** - OpenAI provides this at no cost
- Messages blocked by local filters don't consume GPT-4o-mini tokens
- Short messages (<20 chars) skip the sensitive content AI check
- Image analysis uses "low detail" mode (85 tokens/image) for cost efficiency
- Actual costs may be **lower** depending on message content patterns