

# TruthBot V6 Manual

## INTRODUCTION

TruthBot is a local AI inference and moderation engine built for custom LLM pipelines, plugin integration, and logic-level content reasoning via Honesty Probability (HP) scoring.

## CORE FUNCTION BREAKDOWN

1. LIERWall(prompt\_text: str, context=None) -> dict

Purpose: Runs logical and ethical evaluations.

Returns: {

"sanitized\_prompt": str,

"hp\_score": float,

"notes": str

}

2. score\_hp(prompt: str) -> float

Returns an Honesty Probability score (0.0 - 1.0) for a prompt.

3. apply\_hp\_filter(prompt: str, min\_score=0.5) -> dict

Returns:

{ "hp\_score": float, "prompt": str, "pass": bool }

4. `try_llm(prompt: str, allow_remote=True)`

Routes prompt to either local or external LLM fallback.

5. `detect_capabilities()`

Detects available hardware (GPU/CPU + Torch details).

6. `load_plugins()`

Auto-loads all `.py`` modules in `/llm_plugins/``.

7. `run_diagnostic_pass(source_code)`

Performs a syntax validation and reports issues.

8. GUI class: `TruthBotGUI` (`truthbot_gui.py`)

Tkinter GUI with prompt input and plugin output.

-----

## PLUGIN ROUTING SYSTEM

-----

All plugin files inside `/truthbot/llm_plugins/`` with a `.py`` extension are automatically loaded.

Plugin Example (`math_plugin.py`):

```
```python
```

```
def solve_math(prompt):
```

```
    # custom logic here
```

```
```
```

These are dynamically invoked via plugin manager:

```
```python

plugins = load_plugins()

response = plugins['math_plugin'].solve_math(prompt)

```
```

Default plugins:

- sentiment\_plugin.py
- math\_plugin.py
- qa\_plugin.py

---

## HP SCORE SYSTEM EXPLANATION

---

- score\_hp(): Assigns HP score between 0.0 - 1.0.
- Factors: negation terms, contradictions, length, clarity.
- Children/elderly filters can reject prompts below thresholds.
- Output routing may respect HP score rules.

---

## CLI USAGE GUIDE

---

Run TruthBot via CLI:

```
```bash

python launch_truthbot.py
```

'''

For GUI mode:

```
```bash
```

```
python truthbot_gui.py
```

'''

To check system hardware:

```
```bash
```

```
python -c "from truthbot.hardware import detect_capabilities; print(detect_capabilities())"
```

'''

-----

## GLOSSARY OF INTERNAL CONCEPTS

-----

- LIERWall: Logic/Inference/Ethical Reasoning Wall.
- HP Score: Metric for honesty/clarity of prompt.
- Plugins: Modular logic files for answering or processing.
- Fallback: If local fails, remote LLM is used.
- Diagnostic Pass: Syntax scanner for secure integration.
- Prompt Sanitization: Minor cleanup, not censorship.

-----

## FOLDER STRUCTURE

-----

- /truthbot/

core.py

gui.py

plugin\_manager.py

llm\_interface.py

li\_hp\_scoring.py

hardware.py

llm\_plugins/

math\_plugin.py

sentiment\_plugin.py

qa\_plugin.py

-----