

# AI Media Fake News Investigator – Fake News Detector

---

Turing College – AI Literacy Final Project

---

## Project Overview

This project presents a **single expert-level AI assistant** built with the **Nexos.ai** platform.

The assistant - **AI Media Fake News Investigator** - analyses online information to detect **fake news, bias, propaganda, and harmful narratives**.

It evolved from an earlier multi-agent setup (Fact Checker, Bias Analyst, Propaganda Detector, Aggregator) into one refined, self-contained model.

The design focuses on reliable scoring, ethical operation, and concise, explainable outputs.

---

## Assistant Description

Role: AI Media Fake News Investigator

**Purpose:** Examine text or web links for factual accuracy, neutrality, propaganda intent, safety, and negative influence.

**Platform:** Nexos.ai

**Models tested:** GPT-5, Claude Opus, Gemini 2.5 Pro, Llama

**Output:** Final numeric scores and a short factual summary.

---

## Output Structure

Field	Description	Measurement
<b>Credibility Score (0–100)</b>	Factual correctness.	0 = false / fabricated, 100 = verified
<b>Neutrality Score (0–100)</b>	Degree of balance and absence of bias.	0 = low / 100 = high
<b>Propaganda Score (0–100)</b>	Presence of manipulative or ideological patterns.	0 = low / 100 = high
<b>Safety Score (0–100)</b>	Reliability of source and domain safety.	0 = not safe / 100 = safe
<b>Propaganda Direction</b>	If detected — which side or ideology it supports.	Short description
<b>Negative Influence</b>	Whether the content promotes war, hate, fear, or division.	Short description
<b>Summary</b>	One short, neutral verdict (max ≈ 15 words).	Short description

Field	Description	Mesurement
<b>Sources Checked</b>	Verification references or "no data available."	Short description
<b>Warnings</b>	Notes on unsafe links or unverifiable content.	Short description

## Methodology

1. **Project setup:** Created a new project in the Nexos.ai platform.
2. **Model evaluation:** Opened four parallel chats within the same Nexos project., each using a different model (GPT-5, Claude, Gemini, Llama). Each chat analyzed the same set of links using the assistant's personality prompt. Project creativity was set to analytical for all chats.
3. **Data input:** Uploaded links to five news items representing different content types (satire, verified news, lifestyle, ideological bias, phishing).
4. **Result aggregation:** A fifth chat ("Aggregator Agent") combined outputs into tables for visual and numerical comparison.
5. **Visualization:** Images ([onion\\_news.png](#), [delfi.png](#), [15min.png](#), [fox\\_news.png](#), [netflix.png](#)) were created manually (real screen shots) or with other AI tools to visually represent each analyzed article.
6. **Output documentation:** Results were compiled in Markdown format, providing consistent metrics and summaries for all evaluated links.

## Example Outputs

Example 1 – Satirical / Fake News

### Limbless, Slippery RFK Jr.: ‘Becoming An Eel Is A Sign Of Good Health’



**Input:** [The Onion – “Limbless Slippery RFK Jr.”](#)

Model	Truthfulness Score (0-100)	Bias / Intent	Main Reasoning Summary	Confidence Level	Notes / Anomalies
GPT-5	5	None — satire	Parody content; not factual	High	Correctly flagged as humor
Claude	0	None — satire	Entertainment parody	High	Identified as satire
Gemini	0	Satirical commentary	Known parody; rated high propaganda (45)	Medium	Propaganda score inflated
Llama	10	None — satire	Satire with high neutrality score	Medium	Truthfulness score higher than peers

## Example 2 – Verified Real News



[DELFI INICIATYVA SAUGU](#)

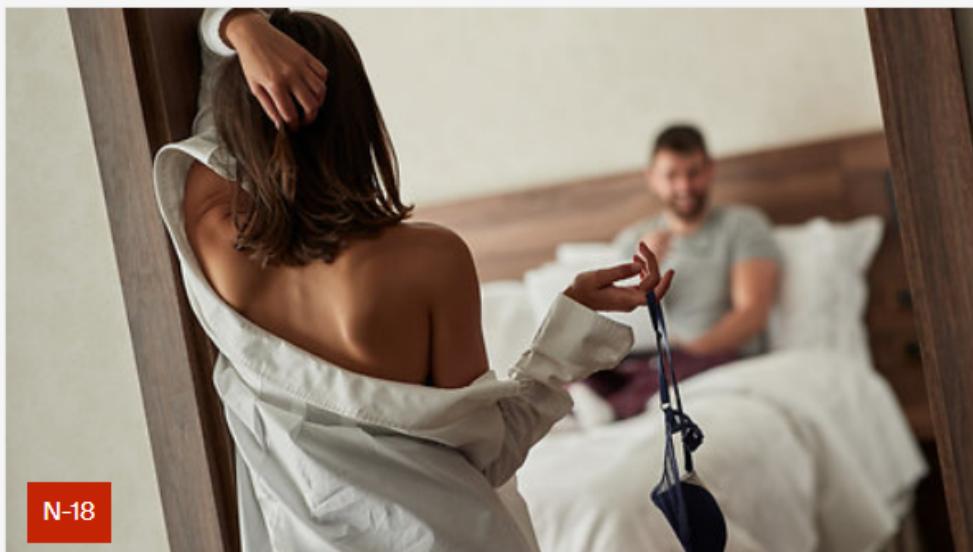
## Lietuviškojo „Taurus“ šarvuočio detektyvas tēsiasi, policija pradėjo ikiteisminį tyrimą (29)

**Input:** Delfi – “Lietuviškojo Taurus šarvuočio detektyvas...”

Model	Truthfulness Score (0-100)	Bias / Intent	Main Reasoning Summary	Confidence Level	Notes / Anomalies
GPT-5	85	None explicit	Credible Lithuanian investigative report	High	Matches reputable source
Claude	85	None detected	Legitimate defense industry news	High	—

Model	Truthfulness Score (0-100)	Bias / Intent	Main Reasoning Summary	Confidence Level	Notes / Anomalies
Gemini	90	N/A	Factual reporting on police case	High	—
Llama	80	None	Legitimate news	High	Slightly lower truthfulness

### Example 3 – Limited Access / Lifestyle Content



**Išsiskyrusi Karina nusimetė tabu – meilę sutiko svingerių klube: atvirai – ką reiškia tokie santykiai**

**Input:** 15min – Article on relationships

Model	Truthfulness Score (0-100)	Bias / Intent	Main Reasoning Summary	Confidence Level	Notes / Anomalies
GPT-5	85	None	Credible lifestyle interview	High	Adult themes warning
Claude	80	None	Alternative relationships personal story	High	—
Gemini	65	None	Subjective interview on sensitive topic	Medium	Lower credibility score
Llama	60	None	Lifestyle piece	Medium	Lowest score among models

### Example 4 – Ideologically Biased News (War Topic)



# Ukraine



**Input:** Fox News – Ukraine conflict coverage

Model	Truthfulness Score (0-100)	Bias / Intent	Main Reasoning Summary	Confidence Level	Notes / Anomalies
GPT-5	70	Pro-Ukraine, critical of Russia	Real news with political framing	High	Emotional tone noted
Claude	65	Right-leaning on Ukraine	Mainstream but partisan framing	High	—
Gemini	65	Right-leaning perspective	Mix of facts and opinion	High	—
Llama	70	Right-leaning	Real news with some bias	High	Propaganda score lower than Gemini

Example 5 – Malicious / Phishing Link



**Input:** [Fake Netflix Login Page](#)

Model	Truthfulness Score (0-100)	Bias / Intent	Main Reasoning Summary	Confidence Level	Notes / Anomalies
GPT-5	0	None — phishing	Malicious credential theft page	High	Extremely unsafe
Claude	0	N/A	Dangerous phishing site	High	Explicit warning
Gemini	0	N/A	Fraudulent phishing attempt	High	—
Llama	0	Criminal/malicious	Unsafe; high propaganda score (90)	Medium	Propaganda score anomalous

## Testing Procedure

1. Use provided prompts to reproduce AI Media Fake News Investigator behaviour.
2. Paste a text or link into Nexus for assistant created.
3. Project creativity - Analytical
4. Wait for the assistant to generate scores and a short summary.

5. Compare fake vs real vs propaganda sources.
  6. Reflect on strengths and weaknesses observed across results.
- 

## Reflection

This task can be done in four-assistant architecture (Fact Checker, Bias Analyst, Propaganda Detector, Aggregator) or can be merged into a single refined model performing all roles simultaneously. This simplifies the workflow, eliminated redundancy, and improves result consistency. The platform itself doesn't offer assistant comparison, everything should be done manually.

### Key learnings

- Prompt design has greater impact than model choice for this task.
  - LLMs can support digital literacy and fact-checking but still need human oversight.
- 

## Supplementary Files

[Prompts](#)

[Gpt5](#)

[Claude-Opus](#)

[Gemini 2.5 Pro](#)

[Llama](#)

[News Links](#)

[Full AI Results Folder](#)

---

**Author:** Sigitas Blechertas

**Date:** October 2025

**Course:** Turing College – AI Literacy Final Project

**Project:** AI Media Fake News Investigator