



Tecnologias de Inteligência Artificial para Detecção de Desinformação e Deepfakes

Equipe de Professores

- Prof. Dr. Ivan Carlos Alcântara de Oliveira (Coordenação) - ivan.oliveira@mackenzie.br
- Prof. Dr. Leandro Zerbinatti - leandro.zerbinatti@mackenzie.br - responsável

MÓDULO 01 | MAPEAMENTO DE TIPOS DE CONTEÚDO



PESQUISADORA
RESPONSÁVEL:
MANUELA



TEXTO

Conteúdo escrito.

[Notícia Falsa]

[Corrente de WhatsApp]

[Postagem em Rede Social]



IMAGEM

Conteúdo visual.

[Foto Antiga]

[Print Falso]

[Imagem IA]

[Gráfico Adulterado]



VÍDEO

Conteúdo manipulado.

[Deepfake]

[Evento Distorcido]

[Legenda Falsa]



ÁUDIO

Foco em apps de mensagem.

[Áudio Falso]

[Voz Clonada por IA]

[Edição/Cortes]



MISTO/MULTIMÍDIA

Combinação para maior credibilidade.

[Texto + Imagem]

[Vídeo + Áudio]

MÓDULO 02 | CRITÉRIOS DE VALIDAÇÃO DE FONTES



PESQUISADOR
RESPONSÁVEL:
ANDRE

Verdade Digital



Pilar Técnico (Infraestrutura Digital)

Registros WHOIS
(idade do domínio)

Certificados SSL/TLS

Reputação/Geolocalização
de IP



Pilar de Autoridade e Transparência

Verificação institucional
(CNPJ/Editorial)

Cruzamento com agências
de fact-checking

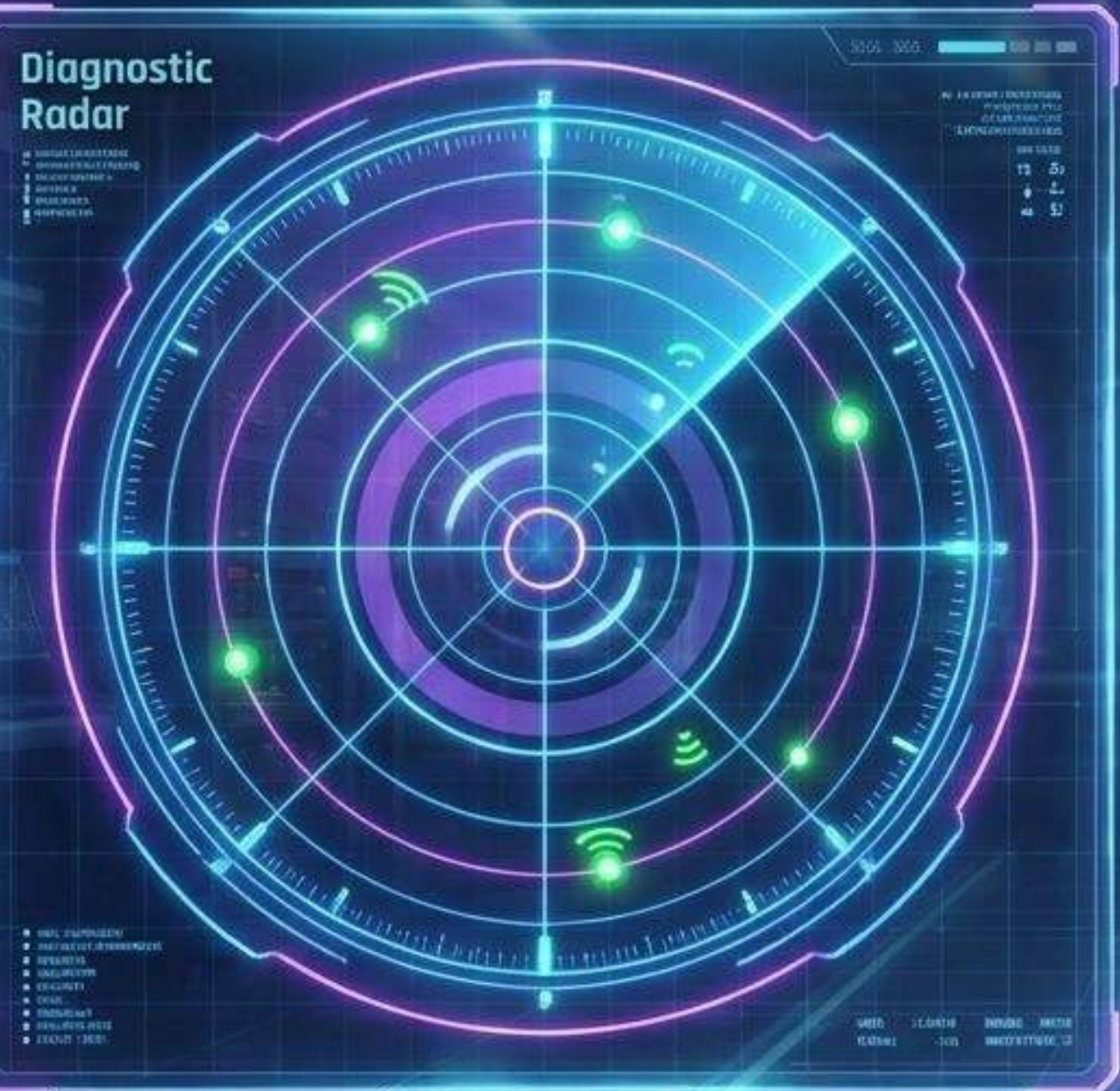
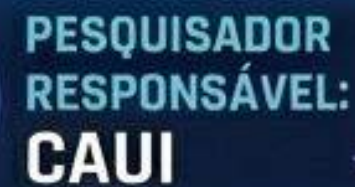


Pilar de Ecossistema e Consenso

Cross-referencing
entre veículos

Rastreabilidade de
fontes científicas

MÓDULO 03 | Tipos de Evidências Exigidas



System Requirements

- ✓ Presença de fontes primárias verificáveis.
- ✓ Confirmação por múltiplas fontes independentes.
- ✓ Metadados e contexto verificáveis (imagens/vídeos).
- ✓ Consistência lógica e factual.
- ✓ Identificação clara de autores e especialistas.
- ✓ Histórico confiável do veículo ou autor.
- ✓ Transparência na apuração.

MÓDULO 04 | Sequência Inicial de Verificação



PESQUISADORA
RESPONSÁVEL:
VITÓRIA



Coleta da
informação

Análise
da fonte

Análise do
conteúdo

Análise
de mídia

Classificação

01

02

03

04

05

Recebimento do
conteúdo;
Identificação do
tipo (texto, imagem,
vídeo).

Quem publicou;
Reputação do
site/perfil;
Histórico;
Autor identificado.

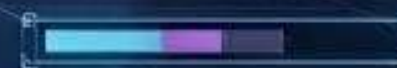
Busca por
inconsistências;
Exageros/sensai-
cionalismo;
Confirmação
cruzada.

Checagem de
contexto/data;
Detecção de
deformações, falhas
e voz robótica.

Nível de
confiabilidade;
Probabilidade de
manipulação;
Veredito final.



OPENTHINK BY
AUTOMATED NEWS EDITOR



MÓDULO 05 | Arsenal de Ferramentas de IA



PESQUISADOR
RESPONSÁVEL:
RAFAEL



Sites e Textos

- **ClaimBuster:** Identifica afirmações suspeitas automaticamente.
- **Google Fact Check Tools:** Busca checagens de agências (falso, errado).



Imagens e Vídeos

- **Deepware Scanner:** Detecção especializada em deepfakes em vídeos.



Pesquisa Nacional

- **Ferramenta UFF:** 94% de precisão na análise de textos do X/Twitter.



Modelos de Linguagem

- **ChatGPT, Claude, Gemini:** Úteis para checar afirmações, mas exigem confirmação em fontes oficiais.



OPINIONES
CONTIDAS SÃO DE RESPONSABILIDADE
DO AUTOR



MÓDULO 06 | O Scorecard de Confiabilidade (Visão Geral)



PESQUISADOR
RESPONSÁVEL:
BRUNO



[RESULTADO FINAL] Exemplo: Acima de 85/100 = Notícia Confiável.
Quanto maior a pontuação, maior a segurança metodológica.

MÓDULO 06 | Detalhamento do Scorecard: Origem e Corroboração



PESQUISADOR
RESPONSÁVEL:
BRUNO

Origem e Fonte (Max: 30 pts)

[+10]

Transparência

- Autor assinado/verificável
- Assinado sem perfil
- Anônimo/Duvidoso

[+5]

[0]

[+10]

Veículo

- Tradicional/Agência (White-list)
- Independente sem histórico ruim
- Banco de desinformação (Black-list)

[+5]

[0]

[+10]

Idade do Domínio

- > 2 anos
- 6 meses a 2 anos
- < 6 meses (descartável)

[+5]

[0]

Corroboração e Evidência (Max: 35 pts)

[+20]

Consenso

- 3+ veículos independentes
- Apenas nicho/regional
- Exclusividade suspeita

[+10]

[0]

[+15]

Fontes Primárias

- Links para docs originais ou vídeos brutos
- Apenas "fontes anônimas" ou opiniões

[0]

MÓDULO 06 | Detalhamento do Scorecard: Texto e Mídia



PESQUISADOR
RESPONSÁVEL:
BRUNO

Análise Textual e Estrutural (Max: 15 pts)

[+5]

Alinhamento Título/Corpo

- Título resume o texto
- Clickbait (promete e não entrega)

[+5]

[0]

[+5]

Sensacionalismo

- Linguagem neutra
- Excesso de CAPS LOCK, exclamações, URGENTE

[+5]

[0]

[+5]

Apelo Emocional

- Foco nos fatos (quem, quando, onde)
- Desenho para gerar raiva/medo via IA

[+5]

[0]

Integridade de Mídia (Max: 20 pts)

[+10]

Contexto / Busca Reversa

- Imagem autêntica do evento
- Foto tirada de contexto (ex: protesto antigo)

[0]

[+10]

Manipulação/IA

- Sem artefatos de edição
- Alta probabilidade de Deepfake não sinalizada

[0]

Módulo 07 | Classificação Final: O Veredito



PESQUISADOR
RESPONSÁVEL:
LUCAS



Verdadeiro

Verificado por fontes
confiáveis e
evidências sólidas.



Falso

Contradiz evidências
verificáveis;
desinformação
confirmada.



Enganoso

Fato real apresentado
fora de contexto ou
distorcido.



Impreciso

Parcialmente correto,
com erros factuais
ou exageros.



Não Verificável

Fontes insuficientes
para confirmar ou
refutar no momento.



Sátira / Opinião

Conteúdo sem
pretensão factual
(humor/opinião).



NOTAS DO MÓDULO 07



MÓDULO 07 | Matriz de Contraste: Entendendo as Nuances



PESQUISADOR
RESPONSÁVEL:
LUCAS

Falso vs. Enganoso



FALSO: Informação 100% inventada.



ENGANOSO: Fato real, mas com contexto distorcido.



Atenção

Deepfakes costumam ser classificados como enganosos quando usam rostos reais em contextos falsos, e não totalmente falsos.

Impreciso vs. Não Verificável



IMPRECISO: Erro identificável na apuração (dato desatualizado, exagero).



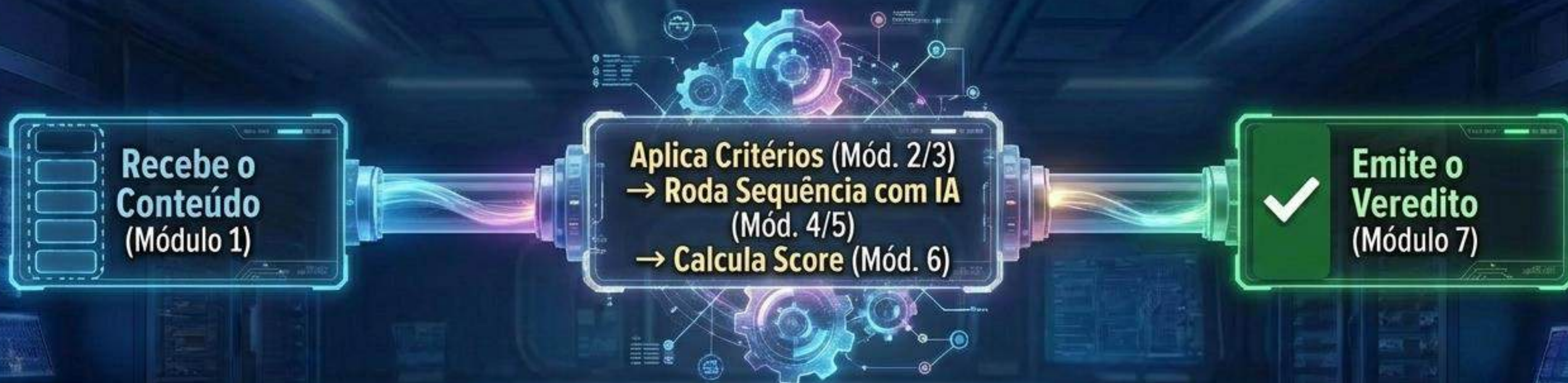
NÃO VERIFICÁVEL: O limbo investigativo. Sem fontes suficientes para confirmar ou refutar a alegação no momento.

A Exceção da Intenção



SÁTIRA / OPINIÃO: Conteúdo humorístico ou opinativo NÃO deve ser classificado como Falso. A intenção factual é ausente por definição. O contexto da publicação dita a regra.

Síntese do Laboratório: O Ciclo de Vida da Checagem



A detecção de desinformação não é intuição; é um **método científico**. Ao combinar **infraestrutura digital**, ferramentas de IA e análise humana crítica, o Projeto Cidadania TecnoMack transforma a verificação em um processo transparente e escalável.



