UNIVALI

UNIVERSIDADE DO VALE DO ITAJAÍ
Computação
Trabalho de Conclusão de Curso (TCC)

PROPOSTA DE TRABALHO

NOME DO ACADÊMICO:	Elian Ferreira
CÓDIGO DE PESSOA:	7321856
E-MAIL DE CONTATO:	ferreira_en@outlook.com
TELEFONE(S) DE CONTATO:	(47) 9 99105-4973

NOME DO PROFESSOR ORIENTADOR:	Felipe Viel
E-MAIL DE CONTATO:	
TELEFONE(S) DE CONTATO:	
DIA DA SEMANA E HORÁRIO DE ATENDIMENTO AO ACADÊMICO:	

CURSO:	**(X)**	**CIÊNCIA DA COMPUTAÇÃO**	**( )**	**ENGENHARIA DE COMPUTAÇÃO**
MODALIDADE DO TRABALHO:	**(X) MONOGRAFIA**		**( ) PRODUTO**	**( ) ARTIGO**
MODALIDADE DA ORIENTAÇÃO:	**(X) PRESENCIAL**		**( ) REMOTA**	
ÁREA DO TRABALHO:	Inteligência Artificial			

Considerando verídicas as Informações fornecidas neste formulário,

encaminhamos a Proposta de Trabalho para avaliação.

ASSINATURA DO ACADÊMICO:

gov.br
Documento assinado digitalmente
ELIAN FERREIRA
Data: 09/12/2025 17:25:06-0300
Verifique em https://validar.iti.gov.br

ASSINATURA DO PROFESSOR ORIENTADOR:

***Sugestão de Banca**:*

ITAJAI (SC), 09/12/2025.


UNIVALI
Escola
Politécnica

APLICAÇÃO DE TÉCNICAS COMPUTACIONAIS PARA DETECÇÃO DE IMAGENS GERADAS POR INTELIGÊNCIA ARTIFICIAL

Elian Ferreira

09 / 2025

Orientador: Felipe Viel, MSc.

Área de Trabalho: Inteligência Artificial

1 INTRODUÇÃO

O avanço da inteligência artificial generativa alterou a criação de conteúdos sintéticos, permitindo que usuários sem conhecimento a programação ou retreinamento de modelos gerem texto, imagens, música e vídeos por meio de interfaces textuais e modelos pré-treinados. O acesso a essas ferramentas se expandiu em áreas como design, marketing e entretenimento. No entanto, surgem os usos indevidos dessas ferramentas, trouxe uma preocupação relacionada a deepfakes, desinformação e questões de propriedade intelectual e de autenticidade de conteúdo (GUILLARO et al., 2025).

Nesse cenário, a verificação forense de imagens se faz essencial na contenção de desinformação e na prevenção da autenticidade digital. Diversas abordagens vêm sendo propostas para diferenciar imagens sintéticas de imagens reais. Estratégias baseadas em análise estatística de entropia permitem detectar irregularidades de distribuição em regiões de baixa complexidade, explorando padrões de aleatoriedade que modelos generativos não reproduzem de forma natural (SUDARSANA et al., 2025). Ao mesmo tempo, métodos de

aprendizado profundo aplicados à geometrias da cena exploram inconsistências em sombras, pontos de fuga e projeção de linhas, demonstrando que modelos generativos ainda não capturam integralmente as propriedades da geometria projetiva (SARKAR et al., 2024).

No domínio estatístico, a lei de Benford, originalmente empregada em auditoria contábil e análise de fraudes, descreve a frequência esperada dos dígitos significativos, em que o dígito 1 ocorre com maior probabilidade que os demais (NIGRINI, 2012; DURTSCHI, HILLISON; PACINI, 2004). Sua aplicação em imagens permite verificar a conformidade das distribuições de intensidade de pixels com as tendências logarítmicas observadas em dados naturais. Desvios expressivos desse padrão podem indicar síntese artificial ou compressão irregular, tornando-a um método complementar aos detectores baseados em aprendizado profundo.

No domínio da frequência, técnicas de fingerprinting espectral analisam desvios em componentes de baixa e alta frequência, permitindo detectar padrões residuais deixados durante o processo de síntese (SHARAFUDDEN; VINOD, 2025; JAMI et al., 2025). Essas abordagens complementam os detectores de aprendizado profundo, oferecendo meios interpretáveis e de baixo custo computacional.

Outra parte da análise concentra-se no impacto de vieses de conjunto de dados. Detectores podem aprender correlações falsas a formato, compressão ou semântica, em vez de características forenses reais. Para reduzir esse problema, foram propostas estratégias de treinamento isentas de vieses, como o paradigma B-Free, que gera imagens sintéticas a partir de reconstruções autocondicionadas de amostras reais, garantindo alinhamento semântico e maior capacidade de generalização a modelos não vistos (GUILLARO et al. 2025).

Em competições internacionais, como a IEE Video and Image Processing Cup 2022, ficou evidente a robustez a novos geradores e a resistência a degradações comuns (compressão, redimensionamento e recorte) são requisitos essenciais para que os detectores funcionem em cenários reais

(COZZOLINO et al., 2022). Além disso, surgiram iniciativas para padronizar e ampliar benchmarks. O So-Fake Dataset incluiu mais de dois milhões de imagens múltiplas categorias a 35 modelos generativos, além de um benchmark fora de distribuição (OOD), permitindo avaliar a capacidade de generalização em contextos de mídia social (HUANG et al., 2025). De forma similar, WILD Dataset foi proposto para atribuição de origem, simulando cenários in-the-wild com geradores comerciais e operações de pós-processamento, o que viabiliza análises de vinculação e atribuição forense (BONGINI et al., 2025).

Apesar desses avanços, trabalhos dizem das limitações de modelos de linguagem multimodais no contexto de detecção de imagens manipuladas. Embora ofereçam explicações interpretáveis, esses modelos ainda não apresentam desempenho comparável a detectores especializados destacando a necessidade de estruturas híbridas ou human-in-the-loop para uso forense (TARIQ et al, 2025). Nesse sentido, a combinação de métodos estatísticos, geométricos e de aprendizado profundo, aliada a bases de dados abrangentes e livres de vieses, constitui um caminho promissor para desenvolver sistemas de verificação que sejam escaláveis e generalizáveis.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Analisar a eficácia e eficiência de técnicas computacionais na detecção de imagens geradas por Inteligência Artificial.

1.1.2 Objetivos Específicos

1. Identificar as principais técnicas forenses utilizadas na detecção de imagens geradas por inteligência artificial;
2. Identificar características geradas por inteligência artificial;
3. Selecionar a metodologia mais adequada para estimar a probabilidade de geração sintética; e

4. Analisar os resultados obtidos com o uso das técnicas forenses selecionadas.

1.1.3 PLANO DE TRABALHO

Para atingir os objetivos propostos, este trabalho será desenvolvido em etapas e subetapas, cada uma voltada à investigação e análise das técnicas forenses aplicadas à detecção de imagens geradas por inteligência artificial. A metodologia adotada abrangerá o levantamento de estudos existentes, a seleção das técnicas, a realização de análises experimentais e a interpretação dos resultados obtidos.

A seguir, são descritas as etapas e atividades que comporão o cronograma deste projeto.

1. Levantamento Bibliográfico: Está etapa irá atender o objetivo 1 e 2 do trabalho e compreende a execução das seguintes atividades.
    a. Definição de critérios: Estabelecer critérios para seleção de trabalhos relacionados á detecção de imagens geradas por inteligência artificial, considerando aspectos como o tipo de abordagem, métricas de avaliação e disponibilidade de dados;
    b. Pesquisa: Realizar buscas para encontrar artigos, ferramentas e produtos similares à análise forense de imagens sintéticas.
    c. Análise: Leitura detalhada dos trabalhos selecionados para então comparar os métodos e técnicas identificadas nos trabalhos selecionados.
    d. Seleção dos métodos: Baseado na análise dos trabalhos relacionados, selecionar as abordagens que serão consideradas na análise comparativa, priorizando aquelas que apresentem resultados replicáveis e documentação técnica acessível; e
    e. Definição dos conjuntos de dados: Identificar e selecionar conjunto de dados públicos que sejam adequados para os

experimentos, levando em conta a relevância e a disponibilidade. Considerando sua utilização em pesquisas recentes e sua relevância para o tema.

2. Implementação e reprodutibilidade: Está etapa irá atender o objetivo específico 2 do trabalho e compreende a execução das seguintes atividades:
    a. Implementação Inicial: Desenvolver uma implementação básica de um método de análise forense utilizando alguma técnica dos trabalhos selecionados.
    b. Reprodutibilidade: Reproduzir as técnicas forenses selecionadas sobre as imagens escolhidas, registrando os resultados e as estimativas de probabilidade de geração sintética; e
    c. Ajustes e Otimizações: Verificar a possibilidade de realizar ajustes e otimizações nos modelos e métodos de análise forense implementados para melhorar o desempenho e a eficiência dos métodos de verificação forense.
3. Análise de resultados: Está etapa irá atender o objetivo específico 3 do trabalho e compreende a execução das seguintes atividades:
    a. Execução dos experimentos: Executar os experimentos com cada uma das técnicas implementadas, registrando as métricas de desempenho definidas na etapa 1;
    b. Coleta e análise de dados: Coletar os resultados de desempenho de cada experimento e realizar uma análise comparativa. Identificar pontos fortes e fracos de cada técnica, discutindo as possíveis causas e implicações dos dados obtidos; e
    c. Relatório de resultados: Elaborar uma análise detalhada apresentando os resultados dos experimentos. Sugerir possíveis melhorias e futuras direções para pesquisas baseadas nos resultados obtidos.

4. Documentação do trabalho: Registro do desenvolvimento da pesquisa com elaboração da monografia com todas as etapas de pesquisa e implementação efetuada.

1.1.4 Cronograma

Apresente o cronograma de execução do seu projeto considerando a execução do TCC II e do TCC III.

Quadro 1. Cronograma de execução do TCC II

Atividade	**12/2025**	**01/2026**	**02/2026**	**03/2026**	**04/2026**	**05/2026**	**06/2026**
1.a. Definição de critérios	XXX_	XX__					
1.b. Pesquisa e Seleção		_XXX	XXX_				
1.c. Análise			__XX	XX__			
1.d. Seleção das técnicas forenses.			_XXX	XXXX	XX__		
1.e. Definição dos conjuntos de dados				__XX	XXX_	_XXX	
2.a. Implementação Inicial					X___	XXXX	XX__
2.b. Reprodutibilidade						__XX	XX__
4. Documentação do trabalho					___X	_XXX	XX__

Quadro 2. Cronograma de execução do TCC III

Atividade	**07/2026**	**08/2026**	**09/2026**	**10/2026**	**11/2026**
2.c. Ajustes e Otimizações	XXX_	XXXX	XX__		
3.a. Execução dos experimentos		XXXX	_XXX	X___	
3b. Coleta e análise dos dados.			__XX	XXXX	
3.c. Relatório de Resultados				_XXX	XX__
4. Documentação do trabalho	___X	___X	__XX	_XXX	XX__

1.2 ANÁLISE DE RISCOS

Caso o seu TCC apresente alguma dependência que ponha em risco o planejamento original, é preciso identificá-la e apresentar um plano alternativo para contornar uma eventual limitação. Apresente essas informações no Quadro 3, no qual as colunas possuem as seguintes definições: (*i*) Risco: descrição do risco identificado; (*ii*) Probabilidade: probabilidade de ocorrer o risco (Alta, Média ou Baixa); (*iii*) Impacto: grau de impacto do risco para o andamento do trabalho (Alto, Médio ou Baixo); (*iv*) Gatilho: evento ou condição que caracteriza a ocorrência do risco; (*v*) Plano de contingência: ações a serem realizadas para contornar os efeitos do risco (é acionado pelo Gatilho).

Quadro 3. Análise de riscos

Risco	Probabilidade	Impacto	Gatilho	Plano de contingência
1. Limitações de Hardware	Média	Média	Lentidão ao processar grandes volumes de dados.	Utilizar serviços de computação em nuvem
2. Falta de dados adequados	Baixa	Média	Inexistência ou insuficiência de dados públicos relevantes	Buscar datasets alternativos ou simular dados sintéticos.

REFERÊNCIAS

BONGINI, Pietro *et al.* **WILD: a new in-the-Wild Image Linkage Dataset for synthetic image attribution**. 2025. Disponível em http://arxiv.org/abs/2504.19595. Acesso em: 11 out. 2025

COZZOLINO, D. et al. *Highlights from the IEEE Video and Image Processing Cup 2022. IEEE Signal Processing Magazine*, v. 39, n. 5, p. 127–132, 2022. Disponível em: https://arxiv.org/abs/2309.12428. Acesso em: 11 out. 2025.

DURTSCHI, C.; HILLISON, W.; PACINI, C. *The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. Journal of Forensic Accounting*, v. 5, p. 17–34, 2004. Disponível em: https://www.researchgate.net/publication/241401706_The_Effective_Use_of_Benford's_Law_to_Assist_in_Detecting_Fraud_in_Accounting_Data. Acesso em: 11 out. 2025.

GUILLARO, A. et al. *A Bias-Free Training Paradigm for More General AI-generated Image Detection.* In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. Disponível em: https://arxiv.org/abs/2412.17671. Acesso em: 11 out. 2025.

HUANG, Y. et al. *So-Fake: Benchmarking and Explaining Social Media Image Forgery Detection.* 2025. Disponível em: https://arxiv.org/abs/2505.18660. Acesso em: 11 out. 2025.

JAMI, N. A. et al. *Patch-Based Deepfake Localization: Unveiling Manipulated Regions in Images through Visual Artifact Analysis. BRAC University Journal of Computer Science*, v. 4, n. 2, p. 45–58, 2025. Disponível em: https://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/26269/20301236%2C20301228%2C21101135%2C20301237_CSE.pdf?sequence=1&isAllowed=y. Acesso em: 11 out. 2025.

NIGRINI, M. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection.* Hoboken, NJ: Wiley, 2012. Disponível em: https://books.google.com.br/books?hl=en&lr=&id=Bh5Vr_I1NZoC&oi=fnd&pg=PR11&dq=info:505QHM_v38kJ:scholar.google.com&ots=qdm8cqysRh&sig=mjqRQTMcq07fp0FJXv1RRBztG8c&redir_esc=y#v=onepage&q&f=false. Acesso em: 11 out. 2025.

SARKAR, S.; GHOSH, S.; ROY, S. *Shadows Don't Lie and Lines Can't Bend: Generative Models Don't Know Projective Geometry.* 2024. Disponível em: https://arxiv.org/abs/2311.17138. Acesso em: 11 out. 2025.

SHARAFUDDEN, S.; VINOD, P. *Dual Residual Learning of Frequency Fingerprints in Detecting Synthesized Biomedical Imagery. Applied Soft Computing*, v. 160, p. 111998, 2025. Disponível em: https://www.sciencedirect.com/science/article/pii/S1568494625002418?casa_token=e5j93NtgARwAAAAA:PU_Cd_E8rFXxhN6zazg91Cpg6tw_L8iBs_Qae2hnMQK8tAiT-SfTOHBWf18rFLWVz3eEoMj7y_Q. Acesso em: 11 out. 2025.

SUDARSANA, K. et al. *LEAD-AI: Lightweight Entropy Analysis for Distinguishing AI-Generated Images from Genuine Photographs. IEEE Transactions on Image Processing*, 2025. Disponível em: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/13480/134800N/LEAD-AI--lightweight-entropy-analysis-for-distinguishing-AI-generated/10.1117/12.3055540.short. Acesso em: 11 out. 2025.

TARIQ, S. et al. *LLMs Are Not Yet Ready for Deepfake Image Detection.* 2025. Disponível em: https://arxiv.org/abs/2506.10474. Acesso em: 11 out. 2025.