



СТАНОВИЩЕ

върху дисертационен труд за придобиване на образователна и научна степен
„доктор“

Автор на дисертационния труд: **маг. инж. Георги Янев Котов**

Тема на дисертационния труд: **„Методи за разграничаване и разпознаване на изображения на човешки лица, генерирани с изкуствен интелект от изображения на реални човешки лица“**

Член на научното жури: **доц. д-р инж. Антония Ташева**

1. Актуалност на разработвания в дисертационния труд проблем

Темата на дисертацията е в безспорно актуалната област за търсене на ефективни и точни методи за разграничаване на реални от ИИ-генерирани или манипулирани изображения, и по-специално човешки лица. Формулираните цел и задачи обхващат в пълнота както научната, така и приложната страна на проблема.

Изследването се фокусира върху изследване на съществуващото състояние на проблема и последващо проектиране на невронни архитектури от вида ResNet-18, Xception, ResNet-50 и ViT-B/16 с PyTorch при стратифицирано разделяне на данните; L^∞ -ограничена адверсариална оценка при пълен достъп с FGSM (Fast Gradient Sign Method) и PGD (Projected Gradient Descent) и финална обяснимост чрез Grad-CAM.

2. Степен на познаване състоянието на проблема и творческа интерпретация на литературния материал.

Дисертационният труд е в обем от 127 страници и цитира общо 123 литературни източника. Източниците са добре подбрани и включват такива с основополагащата теория, както и най-актуалните изследвания в областта. Считам, че докторантът е направил обстоен преглед на състоянието на проблема и задълбочената интерпретация му на литературата са дали стабилна база за постигане на целите в настоящата дисертация.

3. Съответствие на избраната методика на изследване и поставената цел и задачи на дисертационния труд с постигнатите приноси.

Избраната методика на изследване е в съответствие с поставената цел и формулираните задачи на дисертационния труд. Изследователският подход последователно преминава от анализ на съвременното състояние на проблема към разработване и експериментално изследване на хибриден двукласов модел, многопоточна трикласова архитектура и рамка за устойчива и обяснимо откриване (RED - Robust and Explainable Detection), като отделните етапи са методологично обвързани с конкретните задачи на труда. Особено съществено е използването на комбинация от семантични, пространствени и честотни признаци, както и включването на адверсариално обучение, оценка на устойчивостта и средства за обяснимост. По този начин методиката не се ограничава единствено до постигане на висока точност на класификация, а адресира и важни аспекти като устойчивостта и проследимостта на взетите решения.

Заявените научноприложни и приложни приноси произтичат пряко от проведените изследвания и са адекватно подкрепени от получените резултати. Като особено

убедителни следва да се оценят хибридният подход за двукласова детекция, трикласовата многопоточна архитектура с внимание и интегрирането на адверсариална устойчивост и обяснимост в рамката RED, както и последващият им методологичен синтез в RED-MS (Robust and Explainable Detection - многопоточен).

4. Приноси на дисертационния труд

Приемам формулираните в дисертацията 10 приноса. Без да се намалява тяхна стойност, бих препоръчала корекция на тяхната категоризацията. Приносите, означени като „научни“, по своя характер са по-скоро научно-приложни, тъй като не създават нова теория или нов фундаментален научен закон. Тяхната новост е в предлагането, конструирането, интегрирането и експерименталното доказване на нови методи и архитектурни решения за конкретен клас задачи от практически значим научен проблем. Приноси 5 и 6, съответно методологията за сравнителен анализ и метриците имат по-съществен методологичен характер и могат убедително да бъдат определени като научно-приложни. Софтуерната реализация, benchmark тестовите, preprocessing процедурата и Grad-CAM инструментите правилно попадат в категорията приложни приноси.

Научната значимост на приносите се изразява в систематизирането и развитието на методи за детекция на синтетични човешки лица чрез обединяване на различни типове признаци и въвеждане на устойчивост и обяснимост като съвместни критерии. Особено положително следва да се оцени развитието от единичен хибриден модел през многопоточна архитектура и RED до интегрираната RED-MS рамка. Практическата значимост е добре подкрепена от разработената open-source реализация, възпроизводимите benchmark процедури, стандартизираната предобработка и инструментите за експертен преглед.

5. Преценка на публикациите по дисертационния труд:

Основните постижения и резултати от дисертационния труд са публикувани в 4 научни статии (1 самостоятелна и 3 в съавторство, като докторанта е първи автор). Научните статии са представени на международни конференции и са публикувани в реферирани и индексирани издания (Scopus, WoS). Не са известни цитирания на публикациите към настоящия момент.

6. Мнения, препоръки и бележки

Дисертационният труд представлява завършено научно изследване по актуален и значим проблем, свързан с откриването и разграничаването на изображения на човешки лица, генерирани или модифицирани с изкуствен интелект. Разработените модели и методологични рамки са логически свързани с поставената цел и задачи, като са получени убедителни експериментални резултати по отношение на класификационната точност и адверсариалната устойчивост. Положително следва да се оцени и практическата насоченост на разработката, изразена в софтуерна реализация, възпроизводими експериментални процедури и средства за експертен анализ. Приносите са съществени и имат потенциал за развитие и практическо приложение в областта на цифровата форензика и автоматизираното разпознаване на синтетично съдържание.

Като забележки могат да се посочат ограниченото експериментално покритие на някои от заявените аспекти на разработката. По-специално, количествената оценка на обяснимостта и систематичната OOD (out-of-distribution) валидация за обобщаване към невиджани генератори не са представени с пълния обем числени

резултати, предвиден в методологията. Допълнително, оценката на потенциалните демографски пристрастия е идентифицирана като бъдещо направление, но не е реализирана като систематичен експериментален анализ. Тези обстоятелства не намаляват съществено стойността на постигнатите резултати, но следва да бъдат отчетени при формулиране на заключенията относно универсалността, обобщаемостта и безпристрастността на предложените методи.

7. Заключение

Считам, че представеният за рецензиране дисертационен труд отговаря на изискванията на Закона за развитието на академичния състав в република България и му давам обща **положителна** оценка.

Предлагам на докторанта **маг. инж. Георги Котов** да бъде присъдена образователната и научна степен "доктор" по научната специалност „Автоматизирани системи за обработка на информация и управление“ от ПН 5.3 Комуникационна и компютърна техника.

Дата: 08.09.2026 г.

Член на журито:
(доц. д-р инж. Антония Ташева)

OPINION

on a dissertation for the acquisition of the educational and scientific degree "Doctor"

Author of the dissertation: **M.Sc. Eng. Geogri Yanev Kotov**

Topic of the dissertation: **„Methods For Distinguishing And Recognising Ai-Generated Human Face Images From Real Human Face Images“**

Member of the scientific jury: **Assoc. Prof. Eng. Antoniya Tasheva, PhD**

1. Relevance of the problem developed in the dissertation

The topic of the thesis lies within the undeniably topical field of research into effective and accurate methods for distinguishing real images from those generated or manipulated by AI, and in particular human faces. The stated goal and objectives comprehensively cover both the scientific and applied aspects of the problem.

The research focuses on investigating the current state of the art and subsequently designing neural architectures such as ResNet-18, Xception, ResNet-50 and ViT-B/16 using PyTorch with stratified data splitting; L^∞ -constrained adversarial evaluation with full access using FGSM (Fast Gradient Sign Method) and PGD (Projected Gradient Descent), and final explainability via Grad-CAM.

2. Degree of knowledge of the state of the problem and creative interpretation of the literary material

The thesis comprises 127 pages and cites a total of 123 literature resources. The resources are well selected and include both foundational theoretical works and the most recent research in the field. I consider that the PhD candidate has conducted a comprehensive review of the state of the art, and that his in-depth interpretation of the literature has provided a solid foundation for achieving the objectives of this thesis.

3. Compliance of the chosen research methodology and the set goal and objectives of the dissertation with the achieved contributions

The chosen research methodology is well aligned and logically consistent with the stated goal and the objectives set out in the thesis. The research approach moves systematically from an analysis of the current state of the problem to the development and experimental investigation of a hybrid two-class model, a multi-threaded three-class architecture and a framework for robust and explainable detection (RED), with the individual stages being methodologically linked to the specific tasks of the thesis. Of particular significance is the use of a combination of semantic, spatial and frequency features, as well as the incorporation of adversarial training, robustness assessment and explainability tools. In this way, the methodology is not limited solely to achieving high classification accuracy, but also addresses important aspects such as the robustness and traceability of the decisions made.

The stated scientific and applied contributions stem directly from the research conducted and are adequately supported by the results obtained. Particularly convincing are the hybrid approach to two-class detection, the three-class multi-threaded architecture with attention, and the integration of adversarial robustness and explainability into the RED framework, as well as their subsequent methodological synthesis in RED-MS (Robust and Explainable Detection – multi-threaded).

4. Contributions of the dissertation work

I accept the 10 contributions set out in the thesis. Without detracting from their value, I would recommend a revision of their categorization. The contributions labelled as 'scientific' are, by their very nature, more in the realm of applied science, as they do not establish a new theory or a new fundamental scientific law. Their novelty lies in the proposal, construction, integration and experimental validation of new methods and architectural solutions for a specific class of problems within a scientifically significant practical problem. Contributions 5 and 6 – namely, the methodology for comparative analysis and the metrics – are of a more substantial methodological nature and can convincingly be classified as applied scientific. The software implementation, benchmark tests, pre-processing procedure and Grad-CAM tools correctly fall into the category of applied contributions.

The scientific significance of the contributions lies in the systematization and development of methods for detecting synthetic human faces by combining different types of features and introducing robustness and interpretability as joint criteria. The development from a single hybrid model, through a multi-threaded architecture and RED, to the integrated RED-MS framework should be viewed particularly positively. The practical significance is well supported by the developed open-source implementation, the reproducible benchmark procedures, the standardized pre-processing and the expert review tools.

5. Assessment of dissertation publications

The main achievements and findings of the doctoral thesis have been published in four research articles (one authored solely by the doctoral candidate and three co-authored, with the doctoral candidate as the first author). The research articles have been presented at international conferences and published in peer-reviewed and indexed journals (Scopus, WoS). There are currently no known citations of these publications.

6. Opinions, recommendations and notes

This doctoral thesis constitutes a comprehensive piece of scientific research into a topical and significant problem relating to the detection and classification of images of human faces generated or modified using artificial intelligence. The models and methodological frameworks developed are logically linked to the stated aim and objectives, and convincing experimental results have been obtained in terms of classification accuracy and adversarial robustness. The practical focus of the research, manifested in software implementation, reproducible experimental procedures and tools for expert analysis, should also be viewed positively. The contributions are substantial and have potential for further development and practical application in the field of digital forensics and the automated detection of synthetic content.

Among the points to note is the limited experimental coverage of some of the stated aspects of the research. In particular, the quantitative assessment of explainability and the systematic out-of-distribution (OOD) validation for generalization to unseen generators are not presented with the full set of numerical results envisaged in the methodology. Furthermore, the assessment of potential demographic biases has been identified as a future direction but has not been carried out as a systematic experimental analysis. These circumstances do not significantly diminish the value of the results achieved, but should be taken into account when drawing conclusions regarding the universality, generalisability and impartiality of the proposed methods.

7. Conclusion

I believe that the dissertation submitted for review meets the requirements of the Act on the Development of the Academic Staff in the Republic of Bulgaria and I give it an overall positive assessment. I propose that the doctoral student, **M.Sc. Eng. Geogri Kotov** be awarded the educational and scientific degree "Doctor" in the scientific specialty "Automated Information Processing and Management Systems" from the professional field 5.3 Communication and Computer Engineering.

Date: 08.09.2026

Member of the Scientific Jury:

(Assoc. Prof. Eng. Antoniya Tasheva, PhD)