

ФКСТ44-НС1-083

31.08.2026г.



СТАНОВИЩЕ

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

област на висше образование : 5. Технически науки

професионално направление : 5.3. Комуникационна и компютърна техника

Докторска програма : Автоматизирани системи за обработка на информация и управление

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. Ивайло Симеонов Ченчев,
ръководител катедра ИТ, ВУТП**

1. Актуалност на разглеждания в дисертационния труд проблем в научно и научноприложно отношение.

Дисертационният труд разглежда актуален научен и научноприложен проблем, свързан с надеждното разграничаване на реални човешки лица, изображения с разменено лице и изцяло генерирани с изкуствен интелект лица, чиято значимост нараства с развитието на дифузионните модели и свързаните с тях рискове за цифровата идентичност, биометричното удостоверяване и цифровата криминалистика.

Целта е проектиране и оценяване на методи с висока точност като изследването последователно развива хибриден двукласов модел, многопоточна три-класова архитектура, рамката RED и обединената RED-MS архитектура. Поставените задачи обхващат точността, разграничаването на различни типове синтетични изображения, устойчивостта при умишлено внесени смущения в изображенията, обобщаемост към непознати при обучението генеративни модели, което определя темата като актуална както в научен, така и в практически аспект.

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

Докторантът демонстрира добро познаване на съвременното състояние на областта. Анализирани са генеративни модели за синтез на лица - GAN, дифузионни и латентно-дифузионни модели, VAE и техники за размяна на лица - както и основните направления за откриване: CNN, Vision Transformers, хибридни и многопоточни архитектури, пространствени и честотни признаци. Отделено е внимание на атаки и защити при умишлено внесени смущения в изображенията, XAI подходите, наборите от данни и ограниченията на стандартната точност като единствен критерий за оценка.

Литературният обзор служи за извеждане на конкретни изследователски пропуски: загубата на форензична информация при бинарно „real/fake“ класифициране; необходимостта от комбинирание на допълващи пространствени, семантични и спектрални признаци; ограничената устойчивост към въздействия при умишлено внесени смущения в изображенията; и необходимостта от проследима обяснимост. Библиографията съдържа 123 източника и включва фундаментални разработки и

актуални публикации, включително от 2024-2025 г. Това показва достатъчна широчина на литературната основа и адекватно позициониране на собствените разработки.

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

Избраната методика съответства на целта и логиката на дисертационния труд. Тя съчетава литературен и сравнителен анализ, проектиране на невронни архитектури, експериментално обучение и оценяване, адверсариално тестване и методи за обясним ИИ. Методологичният дизайн следва поэтапно усложняване: от двукласов хибриден модел към три-класова многопоточна архитектура, след това към RED и накрая към RED-MS. Използвани са разделяния на данните с балансирано представяне на отделните класове, фиксирани начални стойности, конфигурируеми експерименти и набор от метрики - точност, прецизност, припомняне, F1, MCC, AUC-ROC, ARI и EF.

Експерименталните резултати подкрепят основните твърдения: хибридният модел достига 94.8% точност и превъзхожда разглежданите самостоятелни CNN варианти; многопоточният три-класов модел достига 93.7% обща точност; при RED са отчетени високи референтни резултати за Xception и ResNet-50; а за RED-MS е отчетена точност от 98.14%. Налице е ясна връзка между изследователските въпроси, методите, експериментите и декларираните приноси. Авторът коректно посочва и границите на валидност - статични изображения и ограничен модел на заплахата.

4. Научни и/или научноприложни приноси на дисертационния труд

Анализът на заявените приноси показва известна гранулярност и припокриване, което дава основание десетте формулирани приноса да бъдат обобщени в пет по-ясно разграничени и аргументирани:

Научно-приложни приноси

1. Разработен е хибриден метод за двукласово разграничаване на реални и синтетични изображения на човешки лица, комбиниращ признаци, извлечени чрез конволюционна невронна мрежа, с признаци на ниво пиксел, като експериментално е установено повишаване на точността спрямо самостоятелното им използване.

2. Разработена е многопоточна трикласова архитектура за разграничаване на реални изображения, изображения с разменено лице и изцяло AI-генерирани изображения, която обединява пространствени и честотни признаци чрез механизъм за сливане с внимание.

3. Разработена е рамката RED за оценяване и повишаване на устойчивостта и обяснимостта на моделите за разпознаване на AI-генерирани лица чрез интегриране на адверсариално обучение, оценка на устойчивостта и методи за обяснение на решенията.

4. Разработена е обединена архитектура RED-MS, интегрираща многопоточно трикласово разпознаване, адверсариална устойчивост и обяснимост в единна архитектурна и експериментална рамка.

Приложен принос:

1. Разработена и експериментално апробирана е софтуерна реализация на предложените модели и архитектури, осигуряваща конфигурируемост, възпроизводимост на експериментите и визуализация на основанията за решенията на моделите.

Приносите имат потенциал за приложение при цифрова криминалистика, проверка на медийно съдържание и системи за цифрова идентичност. Същевременно

бих препоръчал при окончателното представяне да се подчертава ясно границата между оригиналната архитектурна и методологична разработка на автора.

5. Преценка на публикациите по дисертационния труд: брой, характер на изданията, в които са отпечатани.

По дисертационния труд са представени четири публикации. Публикацията от Icest 2025 е индексирана в Scopus и Web of Science, а публикацията от COMSCI 2025 – в Scopus. Третата публикация е представена на TechSys 2026 и е посочена като предстояща за публикуване в MDPI Engineering Proceedings (SJR Q3) и индексирание в Scopus и Web of Science. Четвъртата публикация е приета за COMSCI 2026. Тематично публикациите проследяват логично развитието на дисертацията - от хибридно откриване, през многопоточна три-класова архитектура, към устойчивост срещу целенасочени атаки, обяснимост на решенията и обединената многопоточна архитектура.

В предоставените материали не са представени отделни данни за независими цитирания на публикациите от други автори. Поради това не следва да се прави количествен извод за научното отражение чрез цитируемост извън представената публикационна апробация.

6. Мнения, препоръки и бележки

Дисертационният труд е структуриран последователно и представя значителен по обем теоретичен и експериментален материал. Наред с положителните му страни могат да бъдат направени някои бележки, свързани с използваната терминология, езиковото оформление, представянето на илюстративния материал и оформянето на математическите изрази.

В текста на редица места се използват термини и изрази, които представляват буквален или недостатъчно адаптиран превод от английски език и затрудняват четенето и разбирането на изложението. Такива са например „комплементарни форензични сигнали“, което би могло да се замени с „взаимно допълващи се признаци“; „колапсират размяна на лице“ - с „обединяват изображенията с разменено лице“; „форензичен одит“ - с „анализ за установяване на автентичността на изображенията“; „обобщаемостта към невидяни генератори“ - с „обобщаемост към генеративни модели, непознати при обучението“; както и „адверсариалните пертурбации“ - с по-разбираемото „умишлено внесени смущения в изображенията“. Наред с терминологичните неточности се срещат и граматични и синтактични грешки, включително неправилна употреба на предлози, несъгласуване по род и число и неестествени за българския език синтактични конструкции. Честотата на тези неточности показва необходимост от цялостна езикова и терминологична редакция на труда.

Необходимо е и по-прецизно представяне на част от илюстративния материал. При фиг. 4 липсва позоваване на източник, поради което не може еднозначно да се установи нейният произход. Фиг. 8 е представена на английски език, без ясно да е посочено дали представлява авторска разработка, или е заимствана или адаптирана от литературен източник. Препоръчително е произходът на всички фигури да бъде недвусмислено обозначен, включително когато те са разработени от автора. На места се срещат и формулировки, чийто смисъл е трудно разбираем. Такъв пример на стр. 18 е изречението: „Емпиричните глави следват контролирано усложняване на постановката, без смесване на несъвместими протоколи в една таблица.“ От него не става достатъчно ясно какво конкретно се разбира под „контролирано усложняване на постановката“ и „смесване на несъвместими протоколи“. Подобни конструкции следва да бъдат преформулирани с по-ясен и естествен научен език.

По-съществена бележка може да бъде направена по отношение на представянето на метриците и математическия апарат. Като „confusion matrix“ (матрица на

объркванията) е представен математически израз за изчисляване на показател, който по своята същност не представлява матрица на обръкванията (форм. 1.1). При разглежданата трикласова задача матрицата на обръкванията следва да представя съотношението между действителните и предсказаните класове и да позволява анализ на грешките между отделните категории. Налице са и неточности при изписването и оформянето на математическите изрази - срещат се символи без ясно математическо предназначение, като самостоятелно поставени точки, две последователни запетаи и други несъответствия (напр. формули 1.2, 1.3, 1.4 и др.). Тези неточности затрудняват еднозначното интерпретиране на математическите зависимости. Смесването на основни понятия с формули за отделни показатели, заедно с неточностите при математическото им представяне, създава терминологична и методологична неяснота и намалява прецизността при представянето на методологията за оценка на разработените модели.

Посочените бележки не засягат само техническото оформление, тъй като част от тях затрудняват еднозначното разбиране и интерпретиране на изложението. Препоръчвам цялостна езикова, терминологична и техническа редакция на дисертационния труд, включваща използване на утвърдена българска научна терминология, отстраняване на граматичните и синтактичните неточности, прецизиране на математическите означения и формули и еднозначно обозначаване на произхода на използвания илюстративен материал.

7. Заключение

Наред с постигнатите резултати, дисертационният труд съдържа редица терминологични, методологични, граматични, синтактични и редакционни неточности, включително недостатъчно прецизно представяне на някои метрики и математически зависимости, нееднозначно използване на отделни понятия, неточности при оформянето на формули и фигури, както и множество буквално преведени от английски език термини и конструкции, които на места съществено затрудняват четенето и разбирането на изложението.

Посочените забележки не обезценяват получените резултати, но следва да бъдат отчетени при цялостната оценка на качеството и прецизността на представянето на дисертационния труд.

Независимо от направените бележки, считам, че дисертационният труд в своята цялост показва способност на докторанта да формулира изследователски задачи, да разработва и експериментално да оценява подходи за тяхното решаване, и да анализира получените резултати.

Дисертационния труд отговаря на изискванията на Закона за развитие на академичния състав на Република България, на Правилника за неговото приложение, както и на Правилника за условията и реда за придобиване на научни степени в Технически Университет - София. Препоръчвам на уважаемото Научно жури да присъди на **маг. инж. Георги Янев Котов** образователната и научна степен „**доктор**“ в професионално направление 5.3 „Комуникационна и компютърна техника“, докторска програма „Автоматизирани системи за обработка на информация и управление“.

31.08.2026 г.
гр. София

Член на научното жури,
изготвил становището:

/ доц. д-р инж. Ивайло Ченчев /

OPINION

on a dissertation work for the acquisition of an educational and scientific degree “**Doctor**” (PhD)

field of higher education: 5. Technical Sciences

professional field: 5.3. Communication and Computer Technology

Doctoral Program: Automated Information Processing and Management Systems

Author of the dissertation: **Georgi Yanev Kotov, MSc.**

Dissertation title: **“Methods for Distinguishing and Recognizing AI-Generated Human Face Images from Images of Real Human Faces”**

Member of the Scientific Jury: **Assoc. Prof. Ivaylo Simeonov Chenchev, PhD, Eng.,
Head of the Department of Information Technologies,
University of Telecommunications and Post**

1. Relevance of the problem addressed in the dissertation from a scientific and applied-scientific perspective

The dissertation addresses a current scientific and applied research problem: reliably differentiating real human faces from face-swapped images and fully AI-generated faces. This problem is becoming increasingly significant as diffusion models advance and the associated risks to digital identity, biometric authentication, and digital forensics grow.

The aim of the dissertation is to design and evaluate high-accuracy methods. The research progressively develops a hybrid binary classification model, a multi-stream three-class architecture, the RED framework, and the integrated RED-MS architecture. The research tasks cover classification accuracy, differentiation between various types of synthetic images, robustness against intentionally introduced image perturbations, and generalization to generative models unseen during training. These aspects make the topic highly relevant from both scientific and practical perspectives.

2. Degree of knowledge of the state of the problem and creative interpretation of the literature

The doctoral candidate demonstrates a good understanding of the current state of the art in the field. The dissertation analyses generative models for face synthesis—including GANs, diffusion and latent diffusion models, VAEs, and face-swapping techniques—as well as the main detection approaches, such as CNNs, Vision Transformers, hybrid and multi-stream architectures, and spatial and frequency-domain features. It also covers attacks and defense mechanisms involving intentionally introduced image perturbations, XAI approaches, datasets, and the limitations of using standard accuracy as the sole evaluation criterion.

The literature review identifies specific research gaps: the loss of forensic information in binary “real/fake” classification; the need to combine complementary spatial, semantic, and spectral features; limited robustness against intentionally introduced image perturbations; and the need for traceable explainability. The bibliography comprises 123 references and includes both fundamental works and

recent publications, including studies from 2024-2025. This demonstrates sufficient breadth of the literature base and provides an adequate foundation for positioning the candidate's own research contributions.

3. Correspondence of the selected research methodology and the stated aim and objectives of the dissertation with the contributions achieved

The selected methodology is consistent with the dissertation's aim and overall logic. It combines literature review and comparative analysis, neural architecture design, experimental training and evaluation, adversarial testing, and explainable AI methods. The methodological design follows a progressive increase in complexity: from a hybrid binary classification model to a multi-stream three-class architecture, then to RED, and finally to RED-MS. The study uses data splits with balanced class representation, fixed random seeds, configurable experiments, and a set of evaluation metrics, including accuracy, precision, recall, F1-score, MCC, AUC-ROC, ARI, and EF.

The experimental results support the main claims: the hybrid model achieves 94.8% accuracy and outperforms the evaluated standalone CNN variants; the multi-stream three-class model achieves 93.7% overall accuracy; RED reports strong baseline results for Xception and ResNet-50; and RED-MS achieves 98.14% accuracy. The research questions, methods, experiments, and stated contributions are clearly related. The author also appropriately identifies the scope and limitations of the study, including its focus on static images and a limited threat model.

4. Scientific and/or applied-scientific contributions of the dissertation

The analysis of the stated contributions reveals a certain degree of granularity and overlap, which provides grounds for consolidating the ten contributions into five more clearly distinguished and substantiated contributions:

Scientific and Applied Contributions

1. A hybrid method for binary classification of real and synthetic human face images has been developed, combining features extracted by a convolutional neural network with pixel-level features. Experimental results show improved accuracy compared with using either feature type alone.
2. A multi-stream three-class architecture has been developed to distinguish between real images, face-swapped images, and fully AI-generated images, combining spatial and frequency-domain features through an attention-based fusion mechanism.
3. The RED framework has been developed to evaluate and improve the robustness and explainability of models for detecting AI-generated faces by integrating adversarial training, robustness assessment, and methods for explaining model decisions.
4. An integrated RED-MS architecture has been developed, combining multi-stream three-class classification, adversarial robustness, and explainability within a unified architectural and experimental framework.

Applied Contribution:

1. A software implementation of the proposed models and architectures has been developed and experimentally validated, providing experiment configurability, reproducibility, and visualization of the basis for model decisions.
- 2.

The contributions have potential applications in digital forensics, media content verification, and digital identity systems. At the same time, I would recommend that the final presentation clearly distinguish between the author's original architectural and methodological developments and the integration, adaptation, and application of existing methods and technologies.

5. Assessment of the publications related to the dissertation: number and nature of the publications in which they have appeared

Four publications related to the dissertation have been presented. The ICEST 2025 publication is indexed in Scopus and Web of Science, while the COMSCI 2025 publication is indexed in Scopus. The third publication was presented at TechSys 2026 and is indicated as forthcoming in MDPI Engineering Proceedings (SJR Q3), with expected indexing in Scopus and Web of Science. The fourth publication has been accepted for COMSCI 2026. Thematically, the publications follow the dissertation's logical progression—from hybrid detection and a multi-stream three-class architecture to robustness against targeted attacks, explainability of model decisions, and the integrated multi-stream architecture.

The materials provided do not contain separate evidence of independent citations of these publications by other authors. Therefore, no quantitative conclusions regarding the scientific impact of the publications based on citation metrics can be drawn beyond the presented publication record.

6. Opinions, recommendations, and remarks

The dissertation is structured coherently and presents a substantial body of theoretical and experimental material. Alongside its positive aspects, several observations can be made regarding the terminology used, the linguistic presentation, the presentation of illustrative material, and the formatting of mathematical expressions.

In several places, the text uses terms and expressions that appear to be literal or insufficiently adapted translations from English, which hinder the readability and comprehension of the exposition. Examples include “complementary forensic signals,” which could be replaced with “mutually complementary features”; “collapse face swapping” with “group face-swapped images”; “forensic audit” with “analysis for determining image authenticity”; “generalization to unseen generators” with “generalization to generative models not encountered during training”; and “adversarial perturbations” with the more comprehensible “intentionally introduced perturbations in images.” In addition to terminological inaccuracies, grammatical and syntactic errors are also present, including incorrect use of prepositions, lack of agreement in gender and number, and syntactic constructions that are unnatural in Bulgarian. The frequency of these inaccuracies indicates the need for comprehensive linguistic and terminological revision of the dissertation.

Some of the illustrative material also requires more precise presentation. Figure 4 does not include a reference to its source, making it impossible to determine its origin unambiguously. Figure 8 is presented in English, without a clear indication of whether it is the author's own work or has been reproduced or adapted from a literature source. It is recommended that the origin of all figures be explicitly indicated, including cases where they have been created by the author.

Some formulations are also difficult to understand. One such example on page 18 is the sentence: “The empirical chapters follow a controlled increase in the complexity of the experimental setup, without mixing incompatible protocols in a single table.” It is not sufficiently clear what is specifically meant by “controlled increase in the complexity of the experimental setup” and “mixing incompatible protocols.” Such constructions should be reformulated in clearer, more natural academic language.

A more substantial concern relates to the presentation of the evaluation metrics and the mathematical apparatus. A mathematical expression for calculating an indicator is presented as a

“confusion matrix,” although by its nature it does not constitute a confusion matrix (Eq. 1.1). For the three-class classification task considered in the dissertation, a confusion matrix should represent the relationship between the actual and predicted classes and enable analysis of misclassifications across the individual categories. There are also inaccuracies in the notation and formatting of mathematical expressions, including symbols with no clear mathematical purpose, such as isolated full stops, two consecutive commas, and other inconsistencies (e.g., Eqs. 1.2, 1.3, 1.4, etc.). These inaccuracies hinder the unambiguous interpretation of the mathematical relationships. The conflation of fundamental concepts with formulas for individual metrics, together with inaccuracies in their mathematical representation, creates terminological and methodological ambiguity and reduces the precision with which the evaluation methodology for the developed models is presented.

The above observations concern more than the dissertation's technical formatting, since some hinder unambiguous understanding and interpretation of the content. I recommend a comprehensive linguistic, terminological, and technical revision of the dissertation, including the use of established Bulgarian scientific terminology, correction of grammatical and syntactic inaccuracies, refinement of mathematical notation and formulas, and explicit identification of the sources of the illustrative material used.

7. Conclusion

Alongside the results achieved, the dissertation contains a number of terminological, methodological, grammatical, syntactic, and editorial inaccuracies, including insufficiently precise presentation of certain metrics and mathematical relationships, ambiguous use of some concepts, inaccuracies in the formatting of formulas and figures, as well as numerous terms and constructions literally translated from English, which in some places significantly hinder the readability and comprehension of the text.

The above remarks do not diminish the value of the results obtained; however, they should be taken into account in the overall assessment of the quality and precision of the dissertation's presentation.

Despite these observations, I consider that the dissertation, as a whole, demonstrates the doctoral candidate's ability to formulate research problems, develop and experimentally evaluate approaches to solving them, and analyze the results obtained.

The dissertation meets the requirements of the Development of Academic Staff in the Republic of Bulgaria Act, the Regulations for its Implementation, and the Regulations on the Terms and Procedures for Acquiring Academic Degrees at the Technical University of Sofia. I recommend that the esteemed Scientific Jury award **Eng. Georgi Yanev Kotov, M.Sc.**, the educational and scientific degree “**Doctor**” (PhD) in Professional Field 5.3 “Communication and Computer Engineering”, Doctoral Program “Automated Systems for Information Processing and Control”.

31 August 2026
Sofia

Member of the Scientific Jury,
author of the opinion:
/ Assoc. Prof. Ivaylo Chenchev, PhD /