ISSN 2538-8622

INNOVATIONS IN PUBLISHING, PRINTING AND MULTIMEDIA TECHNOLOGIES

INOVACIJOS LEIDYBOS, POLIGRAFIJOS IR MULTIMEDIJOS TECHNOLOGIJOSE

Vol 2 No 1 (2026)

Kaunas, 2026

Contact information
Kauno kolegija Higher Education Institution
Faculty of Informatics, Engineering and Technologies
Department of Informatics and Media Technologies
Pramonės pr. 20, LT-50468 Kaunas, Lithuania
E-mail: daiva.sajek@go.kauko.lt, konferencija.md@go.kauko.lt

TABLE OF CONTENTS

Publishing and Printing

AI for Media, Data Analysis and Cyber Security

Publishing and Printing

INFLUENCE OF PRINTING SUBSTRATE CHARACTERISTICS ON INK TRAPPING BEHAVIOR

Jelena Poljak Gaži, Igor Majnarić, Igor Zjakić, Irena Bates
University of Zagreb, Croatia

Abstract

In offset printing, image reproduction is achieved using four process colors, black, cyan, magenta, and yellow, which are printed sequentially onto the printing substrate. This widely used printing technique enables high-quality and highly detailed image reproduction, and in order to evaluate print quality, several important qualitative parameters are analyzed.

One of the key parameters is ink trapping, which describes the ability of one ink to be accepted over another during the offset printing process. Ink trapping is expressed as a percentage, where higher values indicate better acceptance and adhesion of a subsequent ink layer over a previously printed one. This parameter depends on several interrelated factors that must be carefully balanced in order to achieve optimal print quality. These include ink tack (stickiness), the order in which colors are printed, the speed of the printing machine, and the physical and chemical properties of the printing substrate. Each of these factors can significantly influence the interaction between ink layers during the printing process.

In this study, three specific substrate characteristics were examined in detail: material absorbency, whiteness, and fluorescence of the printing substrate, and their influence on ink trapping behavior. The relationship between ink trapping and substrate characteristics can significantly contribute to the stability, consistency, and overall quality of the printed result. Four different printing substrates were used in the analysis: uncoated substrate, GC1, GC2, and KD board. The comparative analysis of these substrates and their characteristics provides a more comprehensive insight into how ink is accepted and transferred in multi-color offset printing systems, as well as how substrate properties influence final print quality and visual appearance.

Keywords: *Offset printing, ink trapping, absorbency, whiteness, fluorescence*

Introduction

Multicolor offset printing is considered one of the main printing techniques because it enables the reproduction of high-quality printed materials. (Kipphan, 2001). To achieve high-quality prints, a roller system is used that transfers the four process printing colors – black, cyan, magenta, and yellow – onto the printing substrate (Kipphan, 2001). The printing substrates most commonly used in offset printing are coated and uncoated papers, which differ in terms of their absorbency, composition, and brightness of the printing layer.

To define print quality, several qualitative and quantitative parameters are analyzed. One of the main quality parameters is ink trapping. Ink trapping is the ability of one ink layer to accept another ink layer during the printing process and is expressed as a percentage (Nguyen et al., 2022). The higher the percentage, the better the acceptance of the second ink layer over the first printed layer. Ink trapping can be calculated using the Preucil formula, which requires optical density values for CMY inks. Ink trapping is measured in three color areas—red, green, and blue. Red represents the overprint of yellow on magenta, green represents the overprint of yellow on cyan, while blue represents the overprint of magenta on cyan. Ink trapping depends on several parameters, including ink stickiness, ink sequence in the printing press, printing speed, and the characteristics of the printing substrate (Ragab et al., 2012).

The characteristics of the printing substrate that have a significant impact on ink trapping include absorbency, whiteness, and fluorescence. These optical and physical properties directly affect the interaction between printing ink and substrate, as well as the final quality of color reproduction.

Paper absorbency refers to the ability of a material to absorb printing ink. It depends on several key factors, including surface coating, material composition, smoothness, and porosity of the substrate (Sesli et al., 2023). Higher levels of absorbency lead to greater ink penetration into the structure of the substrate, resulting in reduced print density and weaker visual saturation. Therefore, to achieve optimal print quality, it is necessary to select a printing substrate whose absorbency is compatible with the properties of the printing ink used.

Whiteness is an optical property that significantly affects the visual perception of printed colors. It is quantified using a whiteness index. Higher whiteness values contribute to increased contrast and color saturation, while lower values result in less vivid tones (Jurič et al., 2013). For this reason, whiteness is an important parameter in the analysis of print quality and color reproduction.

Fluorescence of the printing substrate refers to the ability of a material to emit visible light after absorbing ultraviolet (UV) radiation. To increase the whiteness of the material, optical brightening agents (OBA) are often added during production (Pasic et al., 2016). The presence of fluorescence and OBAs can affect color perception in prints, especially under different lighting conditions, may cause colors to appear slightly different than expected.

The aim of this paper is to examine how printing substrate characteristics affect colour reproduction, specifically through their impact on ink trapping. The analysis focuses on key substrate properties such as absorbency, whiteness, and fluorescence, and their role in achieving consistent and high-quality print results

Methodology and equipment

Four different printing substrates were used in the study, with their characteristics are listed in Table 1. The printing substrates differ in the type of pulp that is the primary component of each substrate. All printing substrates were printed on the same Koenig & Bauer 105 printing press at a speed of 8,000 sheets per hour under climatic conditions of T = 20–24 °C and RH = 40–60%. HH plate printing plates and D-L and Huber offset water-based CMY printing inks were used. Table 2 presents the optical ink density values of the process inks.

Table 1. Printing substrates – abbreviations and their characteristics

Paper substrate	abbreviations	Grammage (g/m^2)	Composition
Offset paper	off	250	recycled pulp
KD	kd	250	chemical pulp
GC1 paper	gc1	250	mechanical pulp
GC2 paper	gc2	250	mechanical pulp

Table 2. Optical ink density values of process inks

	cyan	magenta	yellow
Optical ink density value	1,38	1,4	1,4

In the printing industry, multi-color printing relies on photomechanical color reproduction, where color tones are produced by halftone areas of process colors (cyan, magenta, yellow, and black). Halftone areas consist of image elements and non-image elements, and the human eye cannot dis-

tinguish the small dots within the image, perceiving only the overall color (Field, 1999).

Technology and control devices in the printing industry are rapidly advancing to achieve the highest quality multi-color prints. One of the main parameters affecting print quality is satisfactory ink acceptance on previously printed ink. Ink trapping is a quality control parameter used to assess ink acceptance behavior on pre-printed ink (Kipphan, 2001). Ink trapping parameters can be calculated by optical measurements using the Preucil method (equation 1), which is based on densitometry.

$$T = \frac{D_{op} - D_1}{D_2} \times 100 \quad (1)$$

Ink trapping parameter (T) is calculated from ink optical density of solid patches: D_1 – ink density of previously printed ink / two inks; D_2 – ink density of last printed ink and D_{op} – ink density of overprint. Whereby, the densities of the ink (D) are observed by using colour filter of the last printed ink.

To compare whether the characteristics of printing substrates affect ink trapping, the following part of the study analyzed the Cobb water absorption parameter and the optical properties of the substrate.

Water absorption parameter refers to measuring the amount of water that paper, board, or corrugated board absorbs over a specific period of time under standardized conditions. The Cobb test, as described in TAPPI T 441, is a simple method for determining the amount of water that paper or board absorbs over a certain period of time. This test defines water absorption as the mass of liquid that one square meter of paper, board, or corrugated board absorbs over a certain period of time under a one centimeter water column (TAPPI, 2009). The calculation of water absorption, or surface area of paper, is performed using equation 2.

$$c(t) = \frac{m_2 - m_1}{\mathrm{P}} \times 10000 \quad (2)$$

where c is Cobb water absorption (g/m²); m_1 is the mass of the sample before testing (g); m_2 is the mass of the sample after testing (g); t is the exposure time (120 sec); and P is the sample surface area exposed to water (100 cm²).

To determine the optical properties of printing substrates, several main optical characteristics were analyzed. Opacity (TAPPI 425 om-06), whiteness (ASTM 313), and fluorescence were measured with an X-Rite eXact spectrophotometer, using standard illuminant D65, a 10° observer, without a polarizing filter, and with an M0 (N0) filter (TAPPI, 2011). Optical measurements were repeated 30 times on each sample.

The degree of opacity is a fundamental property of paper, but its measurement is determined by the reflection ratio. The opacity of a paper is affected by its thickness, the amount and type of filler, the degree of bleaching, the coating, and other factors. Opacity is measured in two steps. First, the sample is placed over a white background with a reflectance of 89%, giving a value defined as R0.89. The second measurement is taken with a black background, which provides nearly zero background reflectance (R0) (Hubbe et al., (2008). Opacity is then calculated using equation 3.

$$Opacity\ (\%) = \frac{R_0}{R_{0.89}} \times 100 \tag{3}$$

Whiteness is a subjectively perceived property based on reflectance data collected across the entire visible spectral range. The CIE Whiteness Index is the global standard for measuring whiteness and is calculated using equation 4.

$$WI = Y + (WI, x)(x_n - x) + (WI, y)(y_n - y) \tag{4}$$

where Y, x, y are the luminance factor and the chromaticity coordinates of the specimen; x_n and y_n are the chromaticity coordinates for the CIE standard illuminant and source used; and WI, x and WI, y are numerical coefficients (ASTM International, 2025).

The colorimetric method, which measures fluorescence, explains the effect of optical brightening agents in paper. A detailed study of the action of these agents was conducted by Allen (Harrison, 1961), in which the optical brightening agent's effect on improving the appearance of paper is explained as producing both a bluing effect (the shade of the paper goes from yellow to blue) and a lightening effect. Standard fluorescence tests were performed at 365 nm in accordance with ISO 2470-2.

Paper fluorescence is rated by the intensity of the reaction under a UV lamp using the Irwin scale, which ranges from "Dead" (0) to "Hybrid" (12). Where dead paper (0) shows no fluorescence; non-fluorescent (1) shows almost no gloss and absorbs almost all light; dull fluorescent (2–3) shows a subtle bluish or whitish glow; weak fluorescent (3–4) is brightly fluorescent; fluorescent (5–6) shows a light bluish white glow; medium fluorescent (7–8) is very bright and intensely bluish white; highly fluorescent (9–10) is intensely bluish white; hybrid (11–12) shows maximum fluorescence (Brixton Chrome, n.d.).

Presentation of research results (Analysis)

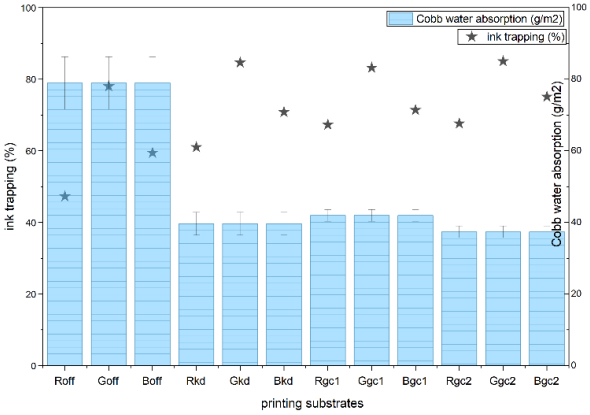

Fig. 1. Correlation between ink trapping and absorption of printing substrates

Figure 1 shows how different printing substrates compare in terms of absorption (bars) and ink trapping efficiency (stars). Uncoated substrate has the highest water absorption values (around 78–80 %), but its ink trapping varies. In contrast, the coated substrates exhibit much lower water absorption (around 37–42%) while generally achieving higher ink trapping percentages, often above 70%. Overall, the results suggest that lower water absorption tends to be associated with better ink trapping performance.

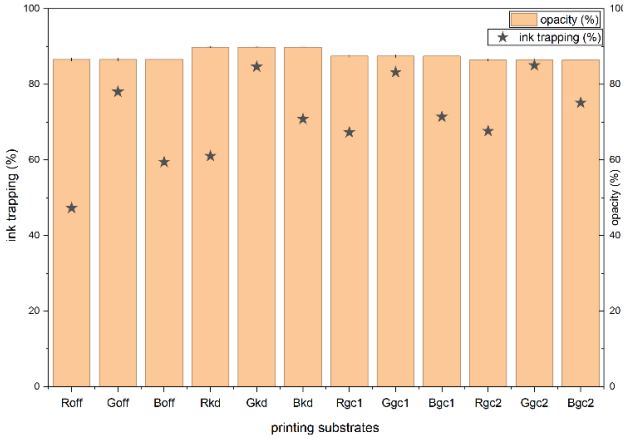

Fig. 2. Correlation between ink trapping and opacity of printing substrates

The results shown in Figure 2 indicate that, although all substrates exhibit high opacity, ink trapping does not change proportionally with opacity. Some substrates with similar opacity values display different ink trapping values, suggesting that opacity alone is not the primary factor influencing ink trapping.

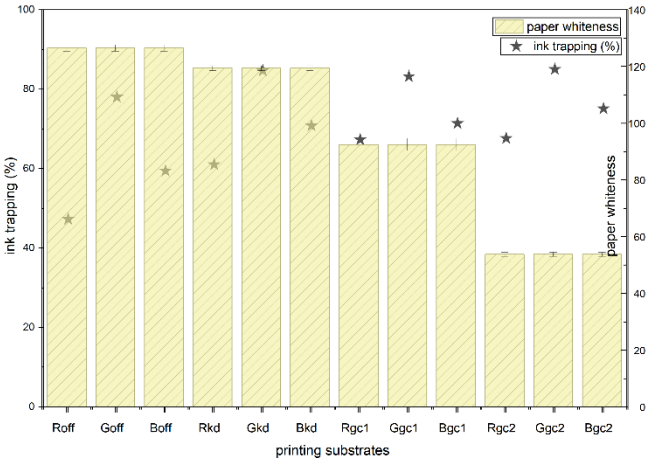

Fig. 3. Correlation between ink trapping and paper whiteness

Results showing the relationship between paper whiteness and ink trapping are presented in Figure 3. A clear trend is observed across the tested substrates: samples with higher whiteness (uncoated substrate, over 90 % whiteness) exhibited lower ink trapping, whereas substrates with moderate whiteness (around 85 %) showed improved trapping performance.

This trend suggests that whiteness itself is not a direct controlling factor for ink trapping. Instead, it is likely correlated with underlying paper properties such as coating composition, surface energy, and ink absorbency, which more directly influence trapping behavior.

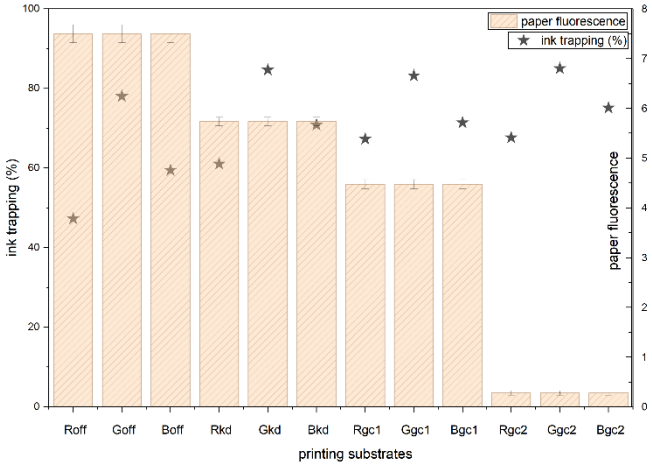

Fig. 4. Correlation between ink trapping and paper fluorescence

Figure 4 compares paper fluorescence with ink trapping across different printing substrates. Uncoated material shows the highest fluorescence (over 90 %), but it ink trapping varies widely, indicating fluorescence alone does not affect trapping performance. Coated materials show moderate fluorescence (around 70 %) with generally consistent ink trapping. The lowest fluorescence appears in the GC2 substrate, but their ink trapping remains relatively high, suggesting that fluorescens does not influence ink trapping performance.

Conclusions

Ink trapping is not determined by a single measurable property but by a combination of surface and structural characteristics of the printing substrate. The results show that optical properties such as opacity do not significantly influence ink trapping, as all tested substrates exhibit similarly high opacity values while showing different trapping performance.

Whiteness and fluorescence also do not directly determine ink trapping behavior. Although variations in these properties are observed between coated and uncoated substrates, the relationship with ink trapping is inconsistent. High whiteness and high fluorescence are associated with both low and variable trapping performance, indicating that these parameters are not controlling factors.

In contrast, water absorption shows a clearer and more consistent relationship with ink trapping. Substrates with lower water absorption generally achieve higher ink trapping values, suggesting that lower porosity contribute to improved interaction between ink layers. Coated substrates, which exhibit lower absorption, consistently demonstrate better trapping performance compared to uncoated materials.

Overall, the results indicate that ink trapping is primarily influenced by surface structure and ink-substrate interaction mechanisms rather than optical characteristics. Key controlling factors are related to porosity, coating presence, and the ability of the substrate to regulate ink penetration and drying behavior.

List of references

1. Kipphan, H. (2001). *Handbook of print media*. Springer. https://doi.org/10.1007/978-3-540-29900-4
2. Nguyen, H. T. K., & Van Huy, P. (2022). *A study of the ink trapping in lithographic printing*. International Journal of Scientific Research in Science and Technology, 16–20. https://doi.org/10.32628/ijsrst2293161
3. Ragab, A. E. R., & Kader, M. a. E. (2012). *Analysis of trapping of color sequences of multicolor offset printing*, 55–61, https://doi.org/10.21608/idj.2012.299562
4. Sesli, Y., Hayta, P., Akgül, A., & Oktav, M. (2023). *Analysis of the parameters determining the effect of coated and uncoated papers on print quality*. 19 Mayıs Sosyal Bilimler Dergisi, 4(3), 130–140. https://doi.org/10.52835/19maysbd.1342475
5. Jurič, I., Karlović, I., Tomić, I., & Novaković, D. (2013). *Optical paper properties and their influence on colour reproduction and perceived print*

quality. Nordic Pulp & Paper Research Journal, 28(2), 264–273. https://doi.org/10.3183/npprj-2013-28-02-p264-273
6. Pasic, R., Kuzmanov, I., & Mijakovska, S. (2016). *Print Quality Control management for papers containing optical brightening agents*. International Journal of Scientific and Engineering Research, 7(2), 271–274. https://doi.org/10.14299/ijser.2016.02.003
7. Field G.G.: Color and Its reproduction (Graphic Arts Technical Foundation, Sewickley, 1999).
8. Technical Association of the Pulp & Paper Industry (TAPPI). (2009). *Water absorptiveness of sized (non-bibulous) paper, paperboard, and corrugated fiberboard (Cobb test*) (TAPPI T 441). https://www.tappi.org
9. Technical Association of the Pulp & Paper Industry (TAPPI). (2011). *Opacity of paper (15/D geometry, illuminant A/2 degrees, 89% reflectance backing and paper backing) (TAPPI T 425)*. https://www.tappi.org
10. Hubbe, M. A. (2008). Paper's appearance: A review. BioResources, 3(2), 627–665. https://doi.org/10.15376/biores.3.2.627-665
11. Hubbe, M.A., Pawlak, J.J., & Koukoulas, A.A. (2008). *Paper's appearance: A review*. BioResources, 3(2), 627–665.
12. ASTM International. (2025). *Standard practice for calculating yellowness and whiteness indices from instrumentally measured color coordinates* (ASTM E313-20). https://doi.org/10.1520/E0313-20R25
13. Harrison, V. G. W. (1961). *Optical properties of paper*. The formation and structure of paper: Transactions of the Second Fundamental Research Symposium, Oxford (pp. 467–485). FRC, Manchester. https://doi.org/10.15376/frc.1961.1.467
14. Brixton Chrome. (n.d.). *Understanding Paper Fluorescence – Advanced version.* https://brixtonchrome.com/pages/understanding-paper-fluorescence-advanced-version#:~:text=It%20will%20not%20have%20the%20brownish%20or,The%20stamp%20on%20the%20right%20is%20DF.

RANDOM FOREST-BASED WASTE PREDICTION IN SUSTAINABLE GRAPHIC PRODUCTION WITHIN INDUSTRY 4.0

Diana Bratić[1], Suzana Pasanec Preprotić[1], Denis Jurečić[1], Jana Žiljak Gršić[2]
[1] University of Zagreb, Croatia
[2] University of Applied Sciences, Croatia

Abstract

Graphic production systems generate significant material waste due to defective prints, registration errors, inefficient material utilization, and suboptimal energy management. In printing processes, waste levels can exceed 10%, contributing to environmental burden and increased operational costs. Within Industry 4.0 environments and ESG-oriented production management, companies require measurable, data-driven mechanisms for resource optimization. The aim of this research is to develop and evaluate a Random Forest-based predictive model for production waste forecasting in sustainable graphic systems using key performance indicators as operational inputs.

A supervised Random Forest regression model was developed to predict waste levels. Because publicly accessible industrial datasets are unavailable due to confidentiality constraints, simulated KPI data was generated to reflect realistic production variability. Input variables include material waste percentage, machine efficiency, number of defective prints, energy consumption, and equipment downtime. The synthetic dataset incorporates controlled variability and inter-variable correlations to approximate real production conditions. The data was divided into training and testing subsets, and cross-validation was applied to ensure model robustness. Model performance was evaluated using mean absolute error, mean squared error, and the coefficient of determination.

The Random Forest model achieved an R^2 value of 0.89, indicating strong explanatory power, with an overall prediction accuracy of 87.32% and a mean absolute error of 3.17%. Simulation scenarios demonstrate substantial waste reduction potential under AI-supported optimization. In the printing process, waste levels decreased from 12.7% to 5.8%, representing a 54.3% reduction. Early detection mechanisms identified 36% of potential waste events prior to occurrence, enabling proactive intervention. Additional simulation results indicate projected improvements in energy

efficiency, defect prevention, and overall cost reduction, estimated at 8%, 12%, and 6%, respectively.
The findings confirm that Random Forest-based predictive modeling can function as an operational tool for ESG-aligned waste management in Industry 4.0 graphic production. KPI-driven monitoring combined with predictive analytics enables early identification of inefficiencies, improved production planning, and systematic reduction of material losses. The model is suitable for integration with sensor systems and production management platforms in smart manufacturing environments, supporting both environmental sustainability and economic performance.

Keywords: *ESG framework, Industry 4.0, key performance indicators, Random Forest, waste prediction.*

Introduction

Graphic production systems are increasingly expected to operate with lower levels of material waste, reduced energy consumption, and a smaller environmental footprint, while still maintaining high quality and cost efficiency. In practice, this is not easy to achieve. Waste in printing and related processes occurs for several reasons, including defective prints, color mismatches, registration errors, and inefficient use of materials and equipment. In many cases, waste levels exceed 10%, which directly increases operational costs and contributes to environmental impact. Under current regulatory and market conditions, such inefficiencies are becoming less acceptable.

The development of Industry 4.0 has introduced new possibilities for improving these processes. Contemporary production environments are supported by interconnected machines, sensor systems, and digital management platforms that enable continuous data collection. This creates a foundation for more precise monitoring and analysis of production performance. Predictive models play an increasingly important role in such environments by enabling data-driven decision-making and improving production efficiency through advanced analytics (Kusiak, 2023). At the same time, ESG requirements are placing additional pressure on companies to measure and report environmental indicators such as waste generation, energy use, and resource efficiency. Key performance indicators represent a fundamental tool for implementing and monitoring sustainability strategies in production systems (Hristov & Chirico, 2019). Machine learning approaches are increasingly used to identify key drivers and improve ESG-related performance indica-

tors in complex operational environments (Dwivedi et al., 2023). In manufacturing systems, key performance indicators also play a central role in measuring sustainability and supporting circular economy strategies (Aljamal et al., 2024), particularly in relation to quality-related indicators that directly influence production efficiency and waste generation (Zuher, 2024). In this context, traditional reactive approaches are no longer sufficient, and there is a clear need for predictive and data-driven solutions.

Artificial intelligence, particularly machine learning, offers a practical way to address these challenges and has been widely recognized as a key enabler of advanced manufacturing systems (Ördek et al., 2024). Its application in smart manufacturing environments enables improved process control, predictive capabilities, and enhanced production efficiency (Rajesh et al., 2022). Predictive models can identify patterns in production data and estimate future outcomes based on current conditions. This makes it possible to detect potential inefficiencies before they lead to actual waste. Among different machine learning methods, Random Forest is often used in similar contexts because it can handle complex relationships between variables and does not rely on strict assumptions about data distribution. It is also relatively robust when working with noisy or heterogeneous data, which is common in production environments.

However, the application of such models in graphic production is limited by the availability of suitable datasets. Industrial data is often confidential, fragmented, or inconsistent, which makes it difficult to use in research. For this reason, this study is based on synthetic data designed to reflect realistic production conditions. The dataset includes controlled variability and correlations between variables, allowing the model to be trained and evaluated in a consistent and reproducible way. While this approach does not replace real production data, it provides a useful framework for testing predictive models.

The main objective of this research is to develop and evaluate a Random Forest-based model for predicting waste levels in graphic production systems. The model uses key performance indicators as input variables, including material waste percentage, machine efficiency, number of defective prints, energy consumption, and equipment downtime. The analysis focuses on whether such a model can support waste reduction and improve resource efficiency within an Industry 4.0 environment aligned with ESG principles. Model performance is evaluated using standard regression metrics, and the results are interpreted in terms of their practical relevance for production systems.

The contribution of this study lies in combining KPI-based monitoring with predictive modeling in the context of graphic production. The results show how machine learning can support earlier detection of inefficiencies and provide a basis for more informed production decisions. In this way, the proposed approach contributes to both operational efficiency and sustainability objectives.

Methodology and equipment

The methodological framework for developing and evaluating the predictive model for waste estimation in graphic production systems combines the definition of relevant operational indicators, the construction of a synthetic dataset that reflects realistic production conditions, and the application of a Random Forest regression model. The objective is to establish a consistent relationship between production parameters and waste generation and to evaluate the predictive capability of the model using standard regression metrics.

Definition of Key Performance Indicators

The predictive model is based on a set of key performance indicators that describe the operational state of the production system. These indicators capture material efficiency, process stability, and production quality, which are the primary sources of waste generation.

Each production instance is represented by a feature vector:

$$x_i = [w_i, e_i, d_i, c_i, t_i] \quad (1)$$

where w_i denotes material waste percentage, e_i machine efficiency, d_i number of defective prints, energy consumption, and t_i equipment downtime.

The target variable represents the total waste level associated with the corresponding production state. For clarity, the selected KPIs are summarized in Table 1.

The selected set of indicators enables a consistent representation of production conditions by integrating material, operational, and quality-related factors into a unified input space. This structure provides the basis for modeling the relationship between process parameters and waste generation in subsequent stages of the analysis.

Table 1. Key Performance Indicators Used as Input Variables.

KPI	Role	Description
Material waste percentage	Share of raw material lost during production	Direct waste indicator
Machine efficiency	Ratio of actual to theoretical output	Process performance
Defective prints	Number of rejected units	Quality-related loss
Energy consumption	Energy usage per production unit	Resource efficiency
Equipment downtime	Duration of production interruptions	Operational instability

Synthetic Data Generation

The predictive model requires a dataset that captures the variability and interdependencies of production parameters in graphic systems. Since real industrial datasets are not publicly available due to confidentiality constraints, a synthetic dataset was constructed to approximate realistic operating conditions. The objective of this approach is to preserve statistical properties of production processes while enabling controlled experimentation and reproducibility. Controlled variability and correlations between variables were introduced to reflect realistic interactions between production parameters under both stable and unstable operating conditions.

Each observation is defined as a realization of the KPI vector under stochastic variability. The generation process assumes that production states fluctuate around a nominal operating point, represented by a mean vector of expected KPI values.

This relationship is modeled as:

$$x_i = \mu + \varepsilon_i \tag{2}$$

where μ denotes the vector of expected values for all KPIs and represents a stochastic deviation from this baseline. The variability component is modeled as a multivariate normal distribution, which is commonly used to represent correlated stochastic processes in multivariate data analysis (Hair et al., 2019):

$$\varepsilon_i \sim \mathcal{N}(0, \Sigma) \tag{3}$$

where Σ is the covariance matrix that encodes dependencies between variables. This formulation allows the dataset to reflect realistic interactions be-

tween production parameters, such as increased defect rates under reduced machine efficiency or higher energy consumption during unstable production phases.

To better approximate real production behavior, the target variable is not treated as independent but is modeled as a function of the input KPIs. The waste level is therefore expressed as:

$$y_i = f(x_i) + \delta_i \tag{4}$$

The function *f* represents an unknown nonlinear mapping between production KPIs and waste generation, which justifies the use of a nonparametric model such as Random Forest. The term δ_i represents an additional noise component capturing unobserved influences and measurement uncertainty.

The dataset includes a wide range of operating conditions by introducing controlled variability across all dimensions of the input space. This ensures that the model is exposed to both stable and unstable production scenarios. The complete dataset is partitioned into training and testing subsets, and k-fold cross-validation is applied to evaluate model generalization and robustness.

This data generation framework provides a consistent and flexible basis for training and validating the predictive model while maintaining realistic assumptions about production variability and system behavior.

Random Forest Model Formulation

The prediction of waste levels is performed using a Random Forest regression model, which is well suited for modeling complex and nonlinear relationships between multiple production variables. Random Forest belongs to a class of tree-based machine learning algorithms that combine multiple decision trees to improve predictive performance and robustness (Sheppard, 2019). Similar applications of process knowledge-based Random Forest regression have shown its suitability for predictive control in nonlinear production processes with multiple operating conditions (Sun et al., 2022).

In the context of this study, the model approximates the functional relationship between the KPI vector and the corresponding waste level defined in the previous section.

The Random Forest model consists of an ensemble of decision trees, where each tree is trained on a bootstrap sample of the dataset. For a given input vector x_i, the predicted waste level is obtained as the average of individual tree predictions:

$$\hat{y}_i = \frac{1}{B}\sum_{b=1}^{B} T_b\ (x_i) \tag{5}$$

where $T_b\ (x_i)$ denotes the prediction of the *b*-th decision tree and B represents the total number of trees in the ensemble. This aggregation reduces variance and improves model stability compared to a single decision tree.

Each decision tree partitions the input space by recursively applying binary splits. At each node, the algorithm selects the optimal split by minimizing the variance of the target variable within the resulting subsets. This is equivalent to minimizing the mean squared error at the node level:

$$\mathrm{MSE}_{node} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{node})^2 \tag{6}$$

where $\hat{y}_{node}$ is the mean value of the target variable within the node.

To introduce diversity among trees and prevent overfitting, Random Forest applies random feature selection at each split. Instead of evaluating all input variables, only a subset of features is considered, which leads to different tree structures and improves generalization performance. Similar approaches combining Random Forest with optimization techniques have been successfully applied in engineering prediction tasks involving complex and nonlinear relationships (Chen et al., 2023; Wengang et al., 2021).

From a regression perspective, the model can be interpreted as an approximation of the conditional expectation of the target variable:

$$\hat{y}_i \approx \mathbb{E}[Y \mid X = x_i] \tag{7}$$

This interpretation is consistent with the formulation introduced in the previous section, where the waste level is modeled as a nonlinear function of the KPI vector. The Random Forest model therefore provides a data-driven approximation of the unknown function $f(x_i)$, enabling accurate prediction of waste levels under varying production conditions. This capability allows the model to capture complex interactions between operational indicators that are difficult to represent using conventional linear approaches. The ensemble structure, combined with bootstrap sampling and feature randomness, makes the model robust to noise and suitable for datasets with complex interdependencies between variables. This is particularly important in graphic production systems, where process variability and measurement uncertainty are inherent characteristics.

Evaluation Metrics

The performance of the proposed model is evaluated using standard regression metrics adapted to the prediction of waste levels in graphic production systems. In this context, y_i represents the observed waste level, while $\hat{y}_i$ denotes the corresponding prediction obtained from the Random Forest model.

The mean absolute error is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{8}$$

This metric provides a direct measure of the average deviation between predicted and actual waste levels and is expressed in the same units as the target variable. It allows for intuitive interpretation of prediction accuracy in terms of percentage deviation.

The mean squared error is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{9}$$

This metric assigns greater weight to larger deviations and is therefore sensitive to extreme prediction errors. It is particularly useful for assessing model stability under conditions with higher variability.

The coefficient of determination is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{10}$$

where $\bar{y}$ denotes the mean observed waste level. This metric quantifies the proportion of variance in waste levels explained by the model and provides an overall measure of predictive performance. In regression analysis, the coefficient of determination is often considered a more informative indicator of model performance compared to alternative error-based metrics (Chicco et al., 2021).

In addition to these metrics, model performance is interpreted in relation to practical thresholds relevant to production systems. Lower error values and higher R^2 indicate that the model successfully captures the relationship between production parameters and waste generation, supporting its applicability in predictive optimization scenarios.

Model Implementation Process

The implementation of the predictive model involves data preparation, model training, parameter configuration, and validation, with the aim of ensuring consistent and reproducible model behavior. Particular emphasis is placed on the model's ability to generalize across different production scenarios. The model implementation and simulation procedures were performed in a Python-based environment using standard machine learning and data analysis libraries suitable for regression modeling and statistical evaluation. The implementation process is structured to ensure stable model performance under varying production conditions and different combinations of operational parameters.

The overall workflow of the proposed model is illustrated in Figure 1.

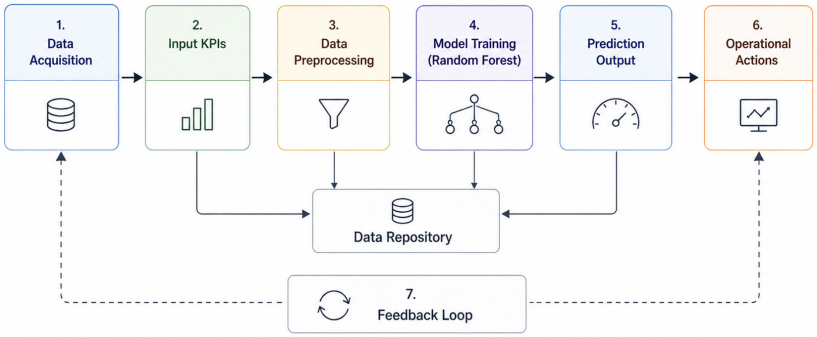

Fig. 1. Workflow of the AI-Based Waste Prediction Model.

The presented workflow summarizes the main stages of the model, from data acquisition and preprocessing to prediction and feedback. Each of these stages is described in detail in the following subsections, with emphasis on their role in ensuring reliable waste prediction and model generalization. The workflow also illustrates the continuous interaction between data analysis, prediction, and operational decision-making within the proposed framework.

Data Preparation

The generated dataset was first prepared for model training by ensuring consistency across all input variables. Since the selected KPIs are expressed in comparable and interpretable units, no additional normalization was required. However, the dataset was inspected to verify the absence of invalid or extreme values that could negatively affect model performance.

The complete dataset was divided into training and testing subsets. The training subset was used to construct the model, while the testing subset was reserved for evaluating predictive performance. This separation ensures that model evaluation is performed on previously unseen data.

Model Training

The Random Forest model was trained using the KPI vectors defined in Section 2 as input features and the corresponding waste levels as the target variable. The training process follows standard supervised learning principles, where the model learns the mapping between input features and target values through repeated exposure to labeled data (Howard & Gugger, 2020). During training, multiple decision trees were constructed using bootstrap sampling, where each tree was trained on a randomly selected subset of observations.

At each split within a tree, a random subset of input features was considered. This mechanism reduces correlation between trees and improves the robustness of the ensemble. The model therefore learns a diverse set of decision rules that capture different aspects of the relationship between production parameters and waste generation.

The training process results in an ensemble model that approximates the underlying nonlinear function $f(x_i)$ introduced in the methodological formulation.

Hyperparameter Configuration

The performance of the Random Forest model depends on several key parameters, including the number of trees, maximum tree depth, and the number of features considered at each split. These parameters were selected to balance model complexity and generalization performance.

A sufficiently large number of trees was used to stabilize predictions and reduce variance. Tree depth was limited to prevent overfitting, ensuring that the model captures general patterns rather than noise in the data. Feature selection at each split was controlled to maintain diversity among trees.

The final parameter configuration was determined based on validation performance, with the goal of achieving stable and consistent predictions across different subsets of the data.

Model Validation

To evaluate model robustness, k-fold cross-validation was applied during the training process. The dataset was partitioned into multiple subsets, and the model was trained and evaluated iteratively, using a different subset as validation data in each iteration.

This approach reduces the dependence of the results on a single data split and provides a more reliable estimate of model performance. The evaluation metrics defined in Section 2 were computed across all validation folds, and the average values were used as the final performance indicators.

The final model was then evaluated on the test dataset to assess its predictive capability under unseen production conditions. This step ensures that the model is not only accurate on training data but also applicable to real-world scenarios.

Results and Discussion

This section presents the results of the proposed Random Forest model and their interpretation in the context of waste prediction, production efficiency, and sustainability. The analysis integrates predictive performance, simulated waste reduction, and system-level implications relevant to Industry 4.0 environments.

Model Performance and Prediction Capability

The predictive model achieved a coefficient of determination of 0.89, indicating that a substantial proportion of the variance in waste levels is explained by the selected KPIs. This result confirms that the relationship between production parameters and waste generation can be effectively captured using a nonlinear ensemble approach. The high explanatory power of the model suggests that the selected operational indicators provide a reliable representation of production behavior and waste formation mechanisms.

The mean absolute error was 3.17%, representing the average deviation between predicted and observed waste levels. This level of error is acceptable in production environments characterized by inherent variability and measurement uncertainty. The mean squared error remained within a stable range, confirming the absence of large deviations and indicating model robustness. Together, these metrics demonstrate that the model maintains stable predictive performance across different production conditions.

The overall prediction accuracy reached 87.32%, defined as the proportion of predictions within an acceptable deviation range. In addition, the model demonstrated the ability to identify 36% of potential waste events before occurrence. This early detection capability enables proactive intervention and represents a shift from reactive quality control toward predictive process management. Such predictive capability is particularly important in production systems where delayed detection can lead to cumulative material and energy losses.

Simulation results further indicate a strong potential for waste reduction across production stages. The most significant improvement was observed in the printing process, where waste levels decreased from 12.7% to 5.8% (Figure 2).

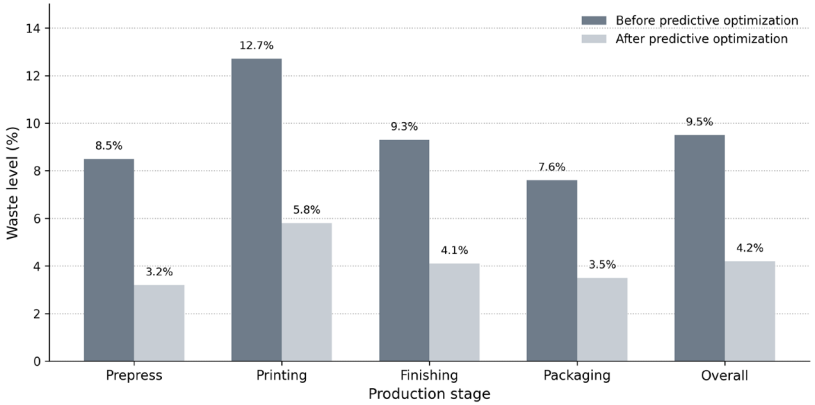

Fig. 2. Comparison of Waste Levels Before and After Predictive Optimization Across Production Stages

Similar trends were observed in prepress, finishing, and packaging, confirming that waste generation is influenced by the overall production system rather than isolated processes.

Operational Impact and System Integration

The application of the predictive model leads to measurable improvements in multiple dimensions of production performance, which is consistent with the use of machine learning techniques for improving sustainability and control in industrial processes (Ahmed & Raihan, 2024). The simulation results indicate an overall waste reduction of approximately 11%, reflecting the direct impact of predictive optimization on material usage. Energy consumption was reduced by 8%, while the number of defective prints decreased by 12%. These improvements resulted in an estimated cost reduction of 6%, demonstrating that waste minimization is closely linked to broader operational efficiency.

Beyond direct prediction, the model provides additional value through pattern recognition. By analyzing production data across multiple observations, the system identifies recurring relationships between KPIs and waste generation. This enables root cause analysis and supports systematic process

adjustments. The continuous comparison between predicted and observed outcomes establishes a feedback mechanism that allows for iterative improvement and adaptation.

The integration of the model into production environments is consistent with Industry 4.0 principles. Data collected from sensor systems can be used as input for real-time prediction, while integration with Manufacturing Execution Systems and Enterprise Resource Planning platforms enables alignment with operational planning and control. Human–AI collaboration remains an essential component, as operators interpret model outputs and implement corrective actions based on domain knowledge.

From a sustainability perspective, the results indicate that predictive waste reduction contributes to reduced environmental impact through lower material consumption and energy use. At the same time, improved efficiency enhances economic performance and supports the transition toward more sustainable and data-driven production systems.

Implementation Challenges and Limitations

Despite the positive results, several challenges must be considered. The use of synthetic data limits the direct applicability of the model to real production environments, as actual data may contain additional variability and complexity not fully captured in the simulation.

Data quality represents a critical constraint. Inconsistent or incomplete production data can reduce model accuracy and require additional preprocessing. The successful implementation of the model therefore depends on reliable data acquisition systems and standardized data management.

Organizational factors also influence adoption. The introduction of AI-based systems may require workforce training and adaptation, as well as acceptance of data-driven decision-making processes.

Finally, system integration presents technical challenges, particularly when connecting legacy equipment with modern Industry 4.0 infrastructures. These limitations indicate that further research is needed to validate the model using real production data and to support its implementation in operational environments.

Conclusion

This study demonstrates that predictive modeling based on key performance indicators can be effectively applied to estimate and reduce waste in graphic production systems. The results confirm that the relationship between production parameters and waste generation can be modeled using a Random Forest approach, even in the absence of real industrial datasets.

The model achieved a high level of predictive performance, with strong explanatory power and acceptable error levels. The simulation results indicate that the application of such a model can lead to substantial reductions in material waste, as well as improvements in energy efficiency, process stability, and overall production performance. In addition to prediction accuracy, the model enables early detection of inefficiencies, which allows for proactive intervention and more consistent production outcomes.

Beyond waste prediction, the results show that the model can support pattern recognition and systematic process optimization. By identifying recurring relationships between production parameters and waste generation, the system provides a basis for continuous improvement and more informed decision-making.

The proposed framework aligns with Industry 4.0 principles by enabling data-driven monitoring, integration with production systems, and support for real-time decision-making. At the same time, it contributes to ESG objectives by reducing environmental impact and improving resource efficiency, while maintaining economic viability.

However, the findings should be interpreted in light of the study's limitations. The use of synthetic data introduces constraints related to realism and generalizability, and further validation using real production data is required. Future research should focus on integrating the model into real production environments, including real-time data acquisition, system interoperability, and adaptive model updating.

Overall, the results indicate that machine learning-based waste prediction has strong potential as a practical tool for optimizing graphic production processes. Its application supports the transition toward more efficient, stable, and sustainable production systems.

List of references

1. Aljamal, D., Salem, A., Khanna, N., & Hegab, H. (2024). Towards sustainable manufacturing: A comprehensive analysis of circular economy key performance indicators in the manufacturing industry. *Sustainable Materials and Technologies, 40*, e00953. https://doi.org/10.1016/j.susmat.2024.e00953
2. Chen, Y., Li, F., Zhou, S., Zhang, X., Zhang, S., Zhang, Q., & Su, Y. (2023). Bayesian optimization based random forest and extreme gradient boosting for the pavement density prediction in GPR detection. *Construction and Building Materials, 387*, 131564. https://doi.org/10.1016/j.conbuildmat.2023.131564

3. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, e623. https://doi.org/10.7717/peerj-cs.623
4. Dwivedi, D., Batra, S., & Pathak, Y. K. (2023). A machine learning based approach to identify key drivers for improving corporate ESG ratings. *Journal of Law and Sustainable Development, 11*(1), 1–15. https://doi.org/10.37497/sdgs.v11i1.242
5. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
6. Howard, J., & Gugger, S. (2020). *Deep learning for coders with fastai and PyTorch*. O'Reilly Media.
7. Hristov, I., & Chirico, A. (2019). The role of sustainability key performance indicators in implementing sustainable strategies. *Sustainability, 11*, 5742. https://doi.org/10.3390/su11205742
8. Kusiak, A. (2023). Predictive models in digital manufacturing: Research, applications, and future outlook. *International Journal of Production Research, 61*(17), 6052–6062. https://doi.org/10.1080/00207543.2022.2122620
9. Ördek, B., Borgianni, Y., & Coatanea, E. (2024). Machine learning-supported manufacturing: A review and directions for future research. *Production & Manufacturing Research*, *12*(1). https://doi.org/10.1080/21693277.2024.2326526
10. Rajesh, A. S., Prabhuswamy, M. S., & Krishnasamy, S. (2022). Smart manufacturing through machine learning: A review, perspective, and future directions to the machining industry. *Journal of Engineering,* 2022, 9735862. https://doi.org/10.1155/2022/9735862
11. Sheppard, C. (2019). *Tree-based machine learning algorithms: Decision trees, random forest and boosting* (2nd ed.). CreateSpace Independent Publishing Platform.
12. Sun, L., Ji, Y., Zhu, X., & Peng, T. (2022). Process knowledge-based random forest regression for model predictive control on a nonlinear production process with multiple working conditions. *Advanced Engineering Informatics, 52*, 101561. https://doi.org/10.1016/j.aei.2022.101561
13. Wengang, Z., Chongzhi, W., Haiyi, Z., Yongqin, L., & Lin, W. (2021). Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers, 12*, 469–477. https://doi.org/10.1016/j.gsf.2020.03.007

14. Zuher, H. A. (2024). Impact quality indicators onto sustainable manufacturing. *Journal of Sustainable Development Innovations, 1*(4), 137–141. https://doi.org/10.61552/JSI.2024.04.001
15. Ahmed, I., & Raihan, A. S. (2024). Machine learning techniques for sustainable industrial process control. In S. K. Paul & S. Kumar (Eds.), *Computational intelligence techniques for sustainable supply chain management* (pp. 141–176). Elsevier.

RESEARCH ON APPROACHES TO THE SELECTION OF MATERIALS IN THE PRODUCTION OF POS PRODUCTS FOR MEDIAMOMENTS PRINT AND PRODUCTION B.V.

Nonna Kulishova[1, 2], Anna Meshcheriakova[2]
[1] Kauno Kolegija Higher Education Institution, Lithuania;
[2] Kharkiv National University of Radio Electronics, Ukraine

Abstract

The article examines the current state and prospects for the development of the publishing and printing industry in the context of constant technological changes and increasing customer requirements. The emphasis is on the need to introduce new materials and improve production processes, which is due to market transformation and the increasing role of visual communications in marketing activities. Particular attention is paid to POS materials (Point of Sale materials), which are an important tool for promoting products and creating brand awareness directly at points of sale. The paper analyzes traditional and modern materials used for the manufacture of POS products. It is noted that cardboard has long remained the main material due to its accessibility, ease of processing and environmental friendliness. At the same time, modern operating conditions for POS materials, in particular increased requirements for durability, moisture resistance, mechanical strength and stability of appearance, limit the possibilities of using cardboard products. In this regard, there is growing interest in the use of polyvinyl chloride (PVC) and other synthetic materials that provide broader functional and design capabilities. The methodological basis of the study is the expert evaluation method, which allowed for a comprehensive comparison of different materials according to technical, economic and environmental criteria. The work defined a list of the main equipment, materials and parameters necessary for conducting a comparative experiment, as well as formulated the evaluation criteria and features of its implementation. This approach ensured the objectivity of the results obtained and took into account practical production conditions. Based on the results of the study, a substantiated choice of material for the manufacture of POS products was made, which most fully meets the specified operating conditions, the specifics of the enterprise's work and its location. It is emphasized that a comprehensive comparison of materials has not only theoretical, but also applied value, since it is focused on current market needs and the possi-

bilities of practical implementation. The conclusions obtained can be used when planning production processes and making decisions on optimizing the material base in the field of manufacturing POS materials.

Keywords: *POS-products, printing materials, cardboard, polyvinyl chloride, marketing, printing.*

Introduction

A wide range of products and a variety of choices have led to the emergence and spread of advertising and information materials at points of sale. Such materials attract the buyer's attention, provide additional information about the product or brand and help make a decision to make a purchase in a retail store filled with similar products.

POS materials (Point Of Sales) areadvertising and information materials placed at the point of sale [1]. POS displays include various advertising elements: displays, shelf design, various designs, packaging, stands, signs, banners, stickers, stands, posters, etc. and include: shelftalkers, shelf organizers, strip holders, placards, posters, show cards, wobblers, brochures, leaflets, flags, three-dimensional designs, dispensers, etc. Among the wide variety of types of POS materials, several categories can be distinguished, according to the time of use: permanent displays (for use longer than six months); semi-permanent displays (term of use – from 2 to 6 months); temporary displays (duration no more than 2 months).

The main tasks of POS materials: attracting the consumer's attention to a specific product, depicting a lifestyle using this product, encouraging buyers to purchase, providing information about the brand. Additional functions of POS materials can also include [1 – 3] localization (as indicating the location of the product); informing about the product; motivating to invite consumers to purchase the product; displaying as emphasizing, drawing attention to the product; branding to form associations with the product or brand.

The study of promotional materials at points of sale has found its place in the works of scholars studying marketing, business and communications. The influence of the quality of POS materials and brand familiarity on trust and intention to purchase a product is noted. This can be relevant for both indoor and online retail, since POS products are an inexpensive but effective way to influence consumer behavior in stores [4, 5].

Usually, POS displays in retail outlets are supplied, placed and serviced by the manufacturer of these POS products. Therefore, manufacturers are interested in manufacturing such products that will deliver advertising messages to consumers as quickly and accurately as possible, and will

perform basic and additional functions within the planned time under the expected conditions of use without loss of quality. All these requirements can be met through the optimal choice of materials from which POS products are made.

Common materials for use in POS are cardboard, paper, polyvinyl chloride, acrylic, banner fabric, aluminum composite sheets, etc.

For the advertising industry and as POS materials, displayboard cardboard is often used – this is a material in the form of a board, which is quite rigid and durable, characterized by the number of layers and thickness. Key advantages of cardboard for POS [3 – 5] include its recyclability, affordable cost and ease of manufacture. Cardboard is easy to print and shape, enabling the production of both flat and three-dimensional structures.

Disadvantages of cardboard include vulnerability to moisture; fire hazard; attractiveness to rodents; fragility and susceptibility to damage during intensive use; the need for additional edge processing, uneven trimming due to the composition of the material.

Polyvinyl chloride (PVC) is colorless, transparent plastic, thermoplastic polymer, a product of vinyl chloride polymerization [6]. Sheet PVC (polyvinyl chloride, vinyl) is known as one of the most durable, versatile, economical and easy-to-process polymer materials available on the market today. PVC is used to create the background of advertising structures, volumetric letters, direct UV printing, the manufacture of stands and POS materials. PVC is easy to manufacture and process, moderately rigid, durable, bends and glues well, and also has a wide palette of thicknesses and colors. In the advertising industry, PVC is often the best choice, due to a number of advantages, including ability to print and form non-flat shapes; light weight, which facilitates transportation and installation; strength, resistance to dents; relatively low cost of PVC; ease of production, cutting and processing and no need for post-printing edge processing; resistance to UV radiation, moisture.

Disadvantages of PVC are low deformation temperature and average frost resistance. As a result, this material is generally unsuitable for large-sized external structures and and requires appropriate recycling.

Thus, PVC is a worthy choice for the advertising industry, due to its relative budget-friendliness, durability, ease of manufacture and suitability for bright printing.

Acrylic, also known as polymethyl methacrylate (PMMA), is a polymer material; it is a product of radical polymerization of methyl methacrylate. It is a thermoplastic material that is often used as an alternative to glass due to its transparency and lightness. Acrylic is the most popular among transpar-

ent and translucent sheet polymer materials. Acrylic is resistant to weathering, can be used both indoors and outdoors for exhibition stands, signs, illuminated letters, light boxes, POS materials.

Acrylic is a good choice for developing high-quality transparent advertising materials and for using them outdoors. However, its high cost and vulnerability to strong impacts/damage significantly limit its use.

Banner fabric is a PVC-coated fabric, which is a type of composite textile material. This material is often used in outdoor advertising, interior design of points of sale and events. Banners are suitable for large-format printing. Such fabric is very durable, flexible, resistant to temperature extremes and ultraviolet radiation, waterproof.

Banner fabric is an interesting and multifunctional solution when it is necessary to use a textile material with resistance to the external environment.

Based on the analysis of the printing materials market, it can be assumed that the use of cardboard materials is not the optimal solution for use in POS products.

There are certain problems associated with the use of cardboard. During production, due to the heterogeneous cellulose structure of cardboard, a poor cut occurs, i.e. defective die-cutting. Due to the multilayer nature of cardboard and its vulnerability to moisture, the material delaminates during operation, which is an important component.

There is an assumption that the transition from the use of cardboard materials to the use of polyvinyl chloride (PVC) in the POS products manufacturing is advisable in view of improving technological and operational characteristics and increasing the durability of finished products.

Therefore, the main of the study is a detailed analysis of the characteristics of the materials used, the most relevant solution for use in POS products and improve the quality of the printed products.

Methodology and equipment

To verify the formulated hypothesis, it is necessary to conduct a study – an expert comparison of the quality of printing materials based on certain criteria, which will form a comprehensive assessment of the copies. This, in turn, will provide a clear understanding of the weaknesses and strengths of each material, reduce the number of production defects and, accordingly, the term and cost of product manufacturing.

A comprehensive assessment of the choice of materials for use in the production of POS materials involves determining a number of criteria for analyzing the choice. It is advisable to form criteria in the following groups: technological, economic, environmental, functional.

The technological criteria employedin the evaluaetion were strength and resistance to mechanical damage; moisture resistance; ease of production; durability of use. Economic benefit (cost) was considered as an economic criterion, recyclability – as environmental, visual appeal – as functional criteria.

The process of manufacturing POS products in publishing and printing is studied at the enterprise Mediamoments Print and Production B.V. (Duivendrecht, Netherlands) [7]. This is a digital printing house specializing in large-format sublimation printing, manufacturing visual communication products on various materials for B2B clients, including retailers, event organizers, advertising agencies, corporate brands and international sporting events.

The study will compare materials for use as a show card or advertising photo booth in the store premises at the point of display of goods.

For POS materials, Mediamoments uses mostly cardboard materials, although polyvinyl chloride, acrylic, banners, self-adhesive stickers, other compositional options (depending on customer requests) are also used. Let's consider the most commonly used materials.

1. Cardboard Katz Displayboard is an environmentally friendly material based on wood pulp, laminated on both sides with white paper. Katz Displayboard offers excellent printing properties, stability, and is used to make tabletop displays, hanging signs, advertising and communication products. The manufacturer offers several types of thickness of such cardboard: 1.6 mm, 2 mm, 3 mm, 5 mm. For use in an outdoor environment, it is necessary to use a special type of Katz Displayboard, which will serve outdoors for up to 12 weeks [8].

During the manufacture of POS products from cardboard Katz Displayboard and its use, a number of shortcomings were identified, namely: low quality of cutting (Fig. 1); low water resistance (Fig. 1); delamination (Fig. 1); low resistance to damage.

Fig.1. Examples of poor die-cutting, delamination, and poor water resistance of Katz Displayboard

Poor quality when material trimming leads to need for re-production or additional operations to improve the product appearance (manual trimming). This, in turn, leads to a cost increase (due to additional material costs). Defects can reach up to 50% of the print run, which increases the time for manufacturing products.

Since cardboard is made from wood pulp, the disadvantages are also related to the material composition (Fig. 1).

2. Kroma Displayboard is a rigid, lightweight cardboard specially designed for indoor use, POS materials and packaging. This durable cardboard is made of cellulose fibers and is 100% recyclable. Kroma Displayboard cardboard appeared on the market quite recently and was offered as a premium version, a higher quality and stable alternative to Katz Displayboard. These two materials are quite similar, but Kroma Displayboard cardboard is significantly more expensive than Katz Displayboard. Kroma is indeed less problematic in production, but still does not have the best cutting quality (Fig. 2).

It is known that recyclability is inherent in cardboard, this material also has a corresponding certificate confirming its environmental friendliness [8].

3. Polyvinyl chloride Vikupor is a foamed PVC sheet for long-term indoor applications, as well as for short-term and medium-term flat applications outdoors (Fig. 3) [8].

Fig. 2. Kroma Displayboard cardboard in the POS products manufacturing

As an alternative material for comparison, polyvinyl chloride (PVC) material, Vikupor was used, which is similar in its physical properties to cardboard, but is more stable in production due to the homogeneity of the material composition and is less susceptible to external factors (Fig. 3). Material shortage is typically limited to 1 copy per 10-piece run, making POS production with Vikupor faster than with cardboard.

Fig. 3. Polyvinyl chloride Vikupor in the POS products manufacturing

PVC is not an environmentally friendly material, but the company manufacturer offers its partial recyclability.

Presentation of research results

The work conducts an experiment to determine the most suitable type of printing material for POS products used in the studied company. Materials for use as a show card or photo stand in retail outlets, indoors, are compared.

Cardboard and polyvinyl chloride were selected as materials similar in terms of application: cardboard Katz Displayboard; polyvinyl chloride Vikupor; cardboard Kroma Displayboard.

These three materials are available for use, are in a similar price category, are popular for use in POS, and have similar physical characteristics (weight, thickness, shape). POS products are manufactured from these materials using the same technology.

Three samples of printing materials were selected for the experiment: cardboard Katz Displayboard, polyvinyl chloride Vikupor, cardboard Kroma Displayboard 3 mm thick.

The supplier [8] provided information on the cost of materials (Table 1).

Table 1. Cost of materials for POS products

Material	Delivery format, mm	Price, EUR/m^2
Cardboard Katz Displayboard	2440×1220×3	8,84
Polyvinyl chloride Vikupor	3050×1560×3	12,76
Cardboard Kroma Displayboard	3050×1560×3	10,40

Nine criteria were selected for comparative assessment: visual appeal; recyclability; moisture resistance; resistance to mechanical damage (damage, scratches, dents); material price; quality of trimming (defects on the edges of the product during trimming); quality of edge processing (paint shedding on the cut, the need for additional actions to improve the appearance of the cuts); durability of use (stability of shape, appearance); resistance to UV radiation.

Experts were asked to rate these criteria when filling out the questionnaire and evaluate which material they thought was the best and worst in nine categories. The assessment was carried out using the ranking method, where the lowest numerical score corresponds to the best result. The survey obtained a rating of the criteria (Table 2).

Table 2. Rating of evaluation criteria

	Evaluation criterion	Criterion rating
K1	Visual appeal	0,022
K2	Recyclability	0,124
K3	Moisture resistance	0,169
K4	Resistance to mechanical damage (damage, scratches, dents)	0,089
K5	Material price	0,138
K6	Trimming quality (defects on the edges of the product when trimming)	0,120
K7	Edge finishing quality (paint shedding on the cut, the need for additional actions to improve the appearance of the cuts)	0,076
K8	Durability of use (stability of shape, appearance)	0,067
K9	Resistance to UV radiation	0,196

To determine the concordance coefficient, the formula is used [9]:

$$W = \frac{12S}{n^2(m^3-m)}, \quad (1)$$

Where n – number of experts; m – number of criteria; S – sum of squared deviation.

$$W = \frac{12 \cdot 1168}{5^2(9^3-9)} = 0{,}779.$$

Based on the result obtained, W=0.779, we can say about a sufficiently high level of consistency of experts' opinions.

Thus, it was found that visual attractiveness is the most important criterion in the production and use of POS materials indoors, and resistance to UV radiation is the least important criterion, which is logical.

In the next step of the experiment, three alternatives were ranked according to nine criteria using expert ratings (alternatives receive ratings from 1 to 3, where 1 is the best option). The degree of consistency of experts' opinions was determined using the concordance coefficients calculated by relation (1). The evaluation results are presented in Table3.

According to the criteria of visual appeal and resistance of materials to UV, the consistency of experts' opinions was mediocre, and according to all other criteria – high, which confirms the reliability of the survey results.

Table 3. Alternatives ratings by criteria

Material			Concordance coefficient
Cardboard Katz Displayboard	Polyvinyl chloride Vikupor	Cardboard Kroma Displayboard	
Visual appeal			
0,47	0,23	0,30	0,52
Recyclability			
0,33	0,50	0,17	1
Moisture resistance			
0,43	0,17	0,40	0,76
Resistance to mechanical damage			
0,47	0,17	0,37	0,84
Material price			
0,40	0,17	0,43	0,76
Trimming quality			
0,50	0,17	0,33	1
Edge finishing quality			
0,50	0,17	0,33	1
Durability of use			
0,40	0,17	0,43	0,76
Resistance to UV radiation			
0,40	0,20	0,40	0,48

The weighting coefficients for the nine criteria were calculated for choosing the best of the proposed alternatives (Table 4).

Table 4. Calculation of the rating weighting factors

Alternatives	Material rating
Cardboard Katz Displayboard	0,42
Polyvinyl chloride Vikupor	0,22
Cardboard Kroma Displayboard	0,36

The results of the alternative comparison are displayed graphically on a radar chart, the best indicator is the lowest weight of the criteria (Fig. 4). It is noticeable that polyvinyl chloride is the leader in comparison with other materials.

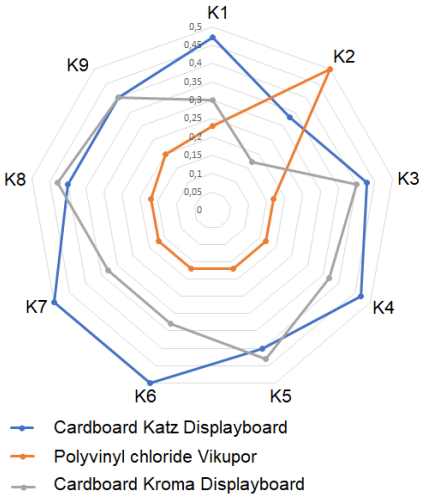

Fig. 4. Comparison of alternatives estimates

Based on the results obtained, it can be concluded that the best material for the manufacture of POS products, namely show cards for use in indoor retail outlets, is polyvinyl chloride.

Its overall rating is better than the rating of two different types of cardboard, according to expert assessments.

It can be concluded that for POS materials, the visual appeal of the product, the quality of edge processing, and durability of use play a decisive role, which really makes sense, because POS materials should primarily attract attention with their neat appearance and pleasant visuals.

It is worth noting that the final results of the choice of materials are also influenced by customer preferences for materials, or the principles/features of the printing company's work.

Conclusions

The study analyzed the work of the enterprise and studied in detail production processes. The main POS materials used in the company was considered, taking into account its specifics of work. The main equipment, materials for research, criteria and features of conducting a comparative experiment were determined.

A review of the studied enterprise and literature made it possible to formulate general recommendations for material choice. The disadvantages of using cardboard for POS materials were identified and an available alternative was proposed.

Based on the research, a material for the manufacture of POS products was selected, which best meets the set conditions, taking into account the specifics of the work and location of the enterprise. A comprehensive comparison was performed taking into account the fact that it has a relevant demand and practical application in today's conditions.

List of references

1. Qader, K. S., Hamza, P. A., Othman, R. N., Anwer, S. A., Hamad, H. A., Gardi, B., & Ibrahim, H. K. (2022). Analyzing different types of advertising and its influence on customer choice. *International Journal of Humanities, Education and Development*, *4*, 8–21.
2. Meshcheriakova, A., & Kulishova, N. (2025). Artificial intelligence as a design tool for a printing shop. In *Memoria del segundo congreso de artes digitales SYNTOPIA* (pp. 20–21). Universidad de Guanajuato.
3. Badawi, B. (2024). The moderating role of POSM (point of sales material) in the relationship between application quality and familiarity on trust and purchase intention. *JBTI: Jurnal Bisnis: Teori dan Implementasi*, *15*(3), 258–272.
4. Brečić, R., Ćorić, D. S., Lučić, A., Gorton, M., & Filipović, J. (2021). Local food sales and point of sale priming: Evidence from a supermarket field experiment. *European Journal of Marketing*, *55*(13), 41–62.
5. Vaičiūtė, D., Gegeckienė, L., & Venytė, I. (2024). Study of the mechanical properties of cardboard with barrier properties. *Innovations in Publishing, Printing and Multimedia Technologies*, 107–114. https://doi.org/10.59476/ilpmt2024.107-114
6. Lewandowski, K., & Skórczewska, K. (2022). A brief review of poly (vinyl chloride) (PVC) recycling. *Polymers*, *14*, Article 3035. https://doi.org/10.3390/polym14153035
7. LinkedIn. (2024). *Mediamoments Print and Production B.V.* https://www.linkedin.com/company/mediamoments/?originalSubdomain=nl

8. Vink Kunststoffen. (2025). *Vink Kunststoffen | Denkt mee. Gaat verder.* https://www.vinkkunststoffen.nl/kunststofsoorten
9. Mendenhall, W., Beaver, R., & Beaver, B. (2020). *Introduction to probability and statistics* (15th ed., Metric version). Cengage Learning.

AI for Media, Data Analysis and Cyber Security

COMPARATIVE ANALYSIS OF EYE-TRACKING AND AI TOOLS FOR PREDICTING VISUAL ATTENTION IN VIRTUAL HUMAN PERCEPTION

Gala Golubović, Sandra Dedijer, Saša Petrović, Sanja Cvetojević, Neda Milić Keresteš, Teodora Gvoka
University of Novi Sad, Serbia

Abstract

Visual attention plays an important role in how users perceive and engage with digital content, particularly in areas such as human–computer interaction and interface design. Eye-tracking is commonly used to study these processes, as it provides detailed and reliable data on visual attention. However, it requires specialized equipment and controlled experimental conditions, which may limit its practical application.

In recent years, artificial intelligence (AI) tools have been developed to predict visual attention based on previously collected eye-tracking data. While these tools are fast and easy to use, it remains unclear how closely their predictions reflect actual human behavior, particularly when observing virtual humans.

This paper presents a comparative study of eye-tracking results and AI-based attention predictions. The study involved 300 participants, whose visual attention was recorded while viewing a set of virtual human stimuli. The same stimuli were then analyzed using the AI tool across several available modes. The comparison was performed using SSIM, MSE, and Pearson correlation, as well as through the analysis of predefined areas of interest.

The results show a strong overall agreement between the two approaches, especially in identifying key facial regions such as the eyes and lips. At the same time, noticeable differences were observed in less prominent areas of the face, as well as during shorter observation periods. AI-generated maps also tend to appear more diffuse compared to eye-tracking data.

These findings suggest that AI tools can provide a useful approximation of visual attention patterns, particularly in the early stages of design. However, they do not fully capture the precision of eye-tracking and should therefore be used as a complementary method rather than a replacement.

Keywords: *eye-tracking, AI predictions, visual attention, virtual human*

Introduction

Understanding visual attention is important in areas such as user interface design, marketing, and human–computer interaction. In the past, researchers have used eye-tracking technology to study it. This method enables precise tracking of eye movements, but requires substantial resources and participant involvement (Duchowski, 2017; Holmqvist, 2011).

As artificial intelligence has advanced, researchers have created attention prediction tools using machine learning and deep learning models trained on eye-tracking data which offer fast and cost-efficient analysis (Kümmerer, Theis & Bethge, 2014; Cornia et al., 2018). However, it is still unclear how reliable they are compared to real eye-tracking data.

This question is especially important when studying virtual humans, since visual attention affects how people perceive emotions and interact with them (Iskra & Tomc, 2016; Calvo & Nummenmaa, 2008).

This study compares eye-tracking results with AI-based attention predictions when people observe virtual humans. The aim is to assess the degree of correspondence between the two methods.

Research Question: Can AI tools reliably reproduce patterns of visual attention obtained through eye-tracking methods?

Hypothesis: There is a significant correspondence between eye-tracking data and AI-based results.

Literature Review

Visual attention is a cognitive process involving the selective distribution of mental resources to relevant stimuli while simultaneously filtering out irrelevant information. In the context of face perception, research shows that the eye and lip regions are the most critical for gathering information about emotions and intentions (Iskra & Tomc, 2016; Calvo & Nummenmaa, 2008; Golubović et al., 2026).

Eye-tracking technology enables the direct measurement of visual attention through the analysis of eye movements, including fixations and saccades, providing objective and precise data on observer behavior (Duchowski, 2017; Holmqvist, 2011). This method is widely applied in fields such as psychology, marketing, and user experience design. In contrast, AI-based attention prediction models rely on machine learning and deep learning algorithms trained with large eye-tracking datasets. These models replicate visual scanning patterns and generate saliency maps indicating probable areas of focus (Kümmerer, Theis & Bethge, 2014; Cornia et al., 2018).

Modern AI tools are accurate and can quickly analyze visual attention without needing participants. However, research shows that these tools can-

not fully replace eye-tracking technology and are mostly used to support it (Lavdas et al., 2021; Lengyel et al., 2021; Šola et al., 2024).

When creating virtual humans, it is important to understand how people pay attention to them. This helps designers make avatars look more realistic and improves the user experience in digital settings.

Methodology

The study is based on a comparative analysis of results obtained through eye-tracking and an AI-based attention prediction tool. The aim was to examine the degree of correspondence between empirically recorded gaze patterns and predictions generated by artificial intelligence models.

Participants and Stimuli

A total of 300 participants took part in the eye-tracking experiment. The stimuli consisted of 12 virtual humans (6 female and 6 male), in which only facial features were varied, while the background and clothing were standardized across all stimuli (Fig. 1).

Participants observed the presented characters and evaluated them according to three parameters: attractiveness, naturalness, and trustworthiness.

Fig. 1. Stimuli

Eye-Tracking Procedure

Data were collected using the GazePoint GP3 eye-tracker. During the experiment, participants' eye movements were recorded, and opacity maps were generated to provide a cumulative visualization of attention distribution.

To analyze the data in more detail, each character's face was split into areas of interest (AOIs). This allowed for a closer look at how attention was spread across different parts of the face (Fig. 2).

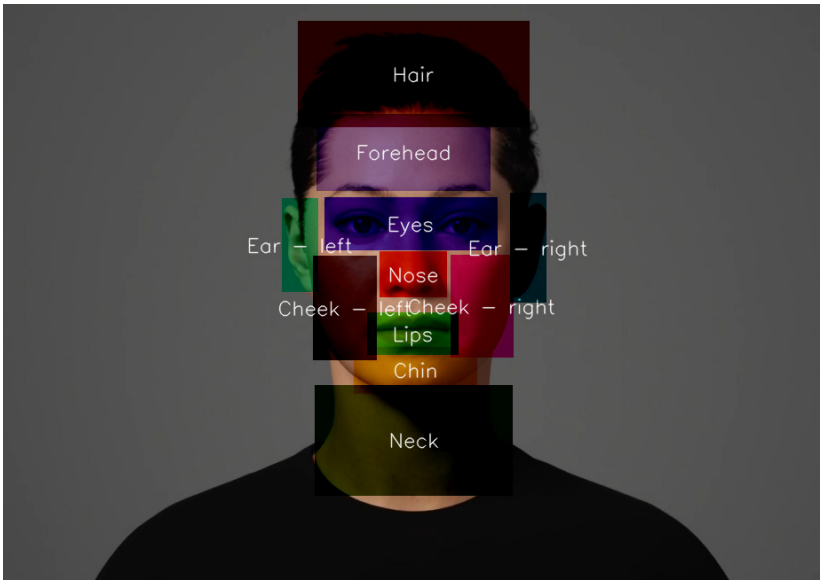

Fig. 2. Areas of Interest generated by GazePoint Analysis Software

For each character, opacity maps were made for three observation times: 3, 4, and 5 seconds. These maps show combined data from all participants (Fig. 3).

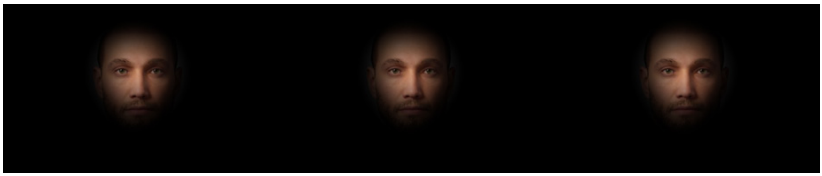

Fig. 3. Opacity maps for times of 3s, 4s, and 5s

AI-Based Attention Analysis

Used AI tool predicted visual attention by generating maps from deep learning models trained on eye-tracking data. The analysis was conducted using the same stimuli as in the eye-tracking experiment, and to capture a broader range of predictions, all available analysis modes offered by the software (*desktop, mobile, marketing material, packaging, poster*, and *shelves*) were used.

Focus maps, which most closely resemble eye-tracking opacity maps, served as the primary basis for comparison (Fig. 4).

Fig. 4. Focus maps in the Attention Insight software

Areas of interest (AOIs) were defined to match those in the eye-tracking analysis, allowing direct comparison of results (Fig. 5).

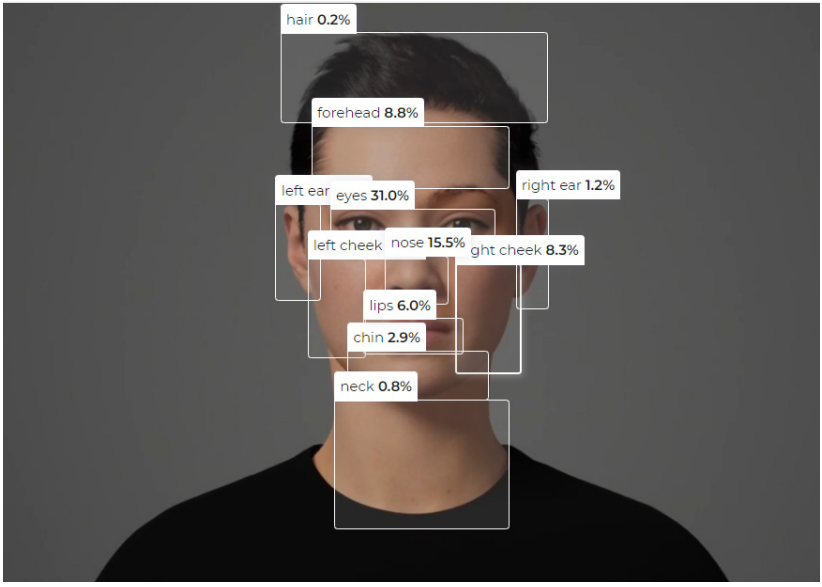

Fig. 5. Areas of Interest generated by AI software

Comparison Metrics

The following metrics were used for the quantitative assessment of correspondence between eye-tracking and AI-generated results:

- SSIM (Structural Similarity Index) – a measure of similarity between two images that accounts for structure, contrast, and luminance. Higher values indicate greater similarity.
- MSE (Mean Squared Error) – represents the average squared difference between pixel values of two images. Lower values indicate smaller error.
- Pearson correlation coefficient – measures the linear relationship between pixel values of two images. Values closer to 1 indicate a stronger positive correlation.

All analyses were conducted in MATLAB software, where comparisons were performed between opacity maps and AI-generated maps across all observation time intervals.

Types of Maps and Representation Selection

Two types of visual attention representations were considered: heat maps and opacity/focus maps. Heat maps provide an intuitive visualization of attention intensity through color gradients (Fig. 6), though their interpretation may differ across software platforms.

Therefore, opacity/focus maps were chosen for comparative analysis, ensuring clearer delineation of areas of interest and more consistent comparison across datasets.

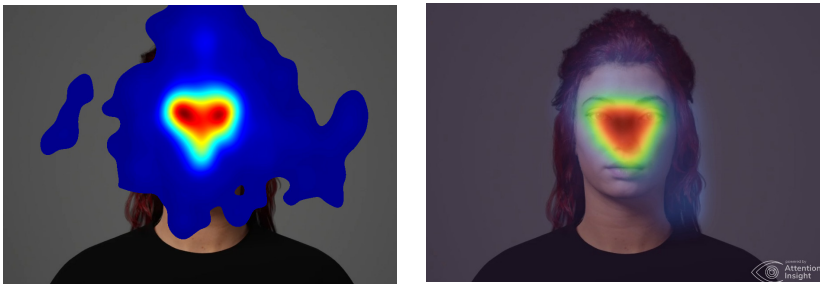

Fig. 6. Heat maps generated by GazePoint Analysis and Attention Insight

Results

The comparison between generated opacity maps and focus maps was conducted in MATLAB software. The analysis involved pairwise comparisons of results from each analysis mode in the AI tool with corresponding results from the GazePoint Analysis software. In total, 180 individual comparison pairs were generated (15 pairs for each of the 12 characters). The analysis included all stimuli and observation intervals (3s, 4s, and 5s) and used quantitative metrics such as SSIM, MSE, and the Pearson correlation coefficient.

Visual analysis revealed a substantial overlap in attention areas between eye-tracking and AI-generated results. AI models successfully identified key areas of interest; however, in some cases, deviations were observed in boundary precision and attention intensity, particularly in peripheral facial regions.

The quantitative analysis indicates a high degree of similarity between maps generated by the two methods:

- SSIM values indicate moderate to high structural similarity, confirming that AI models successfully reproduce global attention patterns.
- MSE values show relatively low error between maps, additionally supporting the quantitative closeness of the results.
- The Pearson correlation coefficient demonstrates a positive and statistically significant relationship between pixel values, indicating consistency in attention distribution across methods.

It was also observed that the degree of correspondence increases with observation time (from 3s to 5s), suggesting that AI models better approximate stabilized attention patterns than the initial phases of visual scanning.

Regarding the analysis modes within AI tool, the *shelves* mode demonstrated the highest level of correspondence, specifically when evaluated with the Structural Similarity Index (SSIM). This suggests that the *shelves* mode produces results most closely aligned with human visual perception according to SSIM. In comparison, other modes such as *mobile* and *packaging* did not achieve as high SSIM values but ranked differently across other metrics.

When evaluated using Mean Squared Error (MSE), the best performance was observed in the *mobile* mode, indicating it had the highest level of objective (mathematical) correspondence. Despite this, *mobile* mode may not appear most similar to human observers. Notably, the *shelves* mode ranked second in MSE-based similarity, demonstrating that it outperformed other modes except mobile mode.

According to the Pearson correlation coefficient, the highest correlation was observed in the *packaging* mode, suggesting the strongest relationship

in pixel value trends for this mode. The lowest correlation was associated with the *marketing material* mode. This suggests that the *packaging* mode shows the strongest correspondence for this metric.

Observation Time

Since both software tools support defining areas of interest (AOIs) and recording the percentage of attention directed to each region, an additional comparison was made using these metrics. As the strongest alignment between eye-tracking results and AI predictions emerged at the 5-second interval, this analysis focused solely on data from that period.

The comparison of these values showed that predictions generated with the *packaging* mode most closely matched recorded gaze behavior. The *shelves* mode was the next most similar (Table 1).

Table 1. Proportion of total observation time allocated to the area of interest

		Gaze-Point Analysis	**Attention Insight**				
Virtual Human	AOI	Ave Time (%)	Desktop (%)	Marketing material (%)	Mobile (%)	Pack-aging (%)	Shelves (%)
	Eyes	41.7	21.9	21.6	21.8	21.3	**26.1**
	Nose	11.9	15.5	25.6	16.4	**9.1**	18.9
	Lips	7.4	8.3	10.1	12.5	6.3	**6.7**
	Hair	4.8	0.0	0.0	0.1	**2.2**	0.0
	Neck	3.2	4.0	**3.1**	2.5	2.7	1.7
	Left ear	1.2	1.4	0.0	0.7	4.3	**1.1**
	Right ear	1.3	**1.8**	0.1	2.0	**1.8**	0.6
	Chin	3.2	10.8	8.3	9.5	6.6	**5.9**
	Left cheek	4.4	6.5	**3.5**	6.5	6.9	9.0
	Right cheek	2.6	6.9	**5.9**	9.0	6.6	7.9
	Forehead	14.2	7.9	5.6	9.7	**11.7**	9.0
	Eyes	37.3	22.4	**35.4**	22.1	18.5	21.9
	Nose	10.9	16.0	33.1	17.6	**9.4**	16.4
	Lips	5.1	10.5	8.8	10.1	**6.8**	7.3
	Hair	7.2	4.9	0.7	2.9	2.0	**9.1**
	Neck	3.4	**2.9**	0.7	1.0	0.6	0.4
	Left ear	2.3	0.1	0.1	0.9	**2.7**	0.6
	Right ear	1.2	1.0	0.4	2.4	2.5	1.2
	Chin	3.1	5.5	**3.0**	5.0	4.4	3.6
	Left cheek	2.5	4.7	**2.3**	6.2	8.1	5.7
	Right cheek	2.4	5.3	**4.4**	5.2	7.1	6.9
	Forehead	16.0	8.5	9.8	**12.5**	9.6	10.6

	Eyes	37.6	21.5	20.3	22.5	12.8	**22.8**
	Nose	8.0	13.4	28.7	13.8	**10.4**	16.0
	Lips	5.2	8.4	9.1	8.6	7.7	**6.6**
	Hair	8.2	0.3	0.1	1.0	**1.3**	0.5
	Neck	3.3	4.3	0.5	2.4	6.1	**2.6**
	Left ear	1.2	1.5	0.0	**1.4**	3.4	2.0
	Right ear	1.7	2.4	0.6	**2.4**	2.5	3.4
	Chin	2.8	7.5	**4.2**	8.8	9.7	4.3
	Left cheek	3.3	5.9	**3.5**	6.6	5.8	6.7
	Right cheek	3.3	7.2	**6.8**	8.5	6.8	10.5
	Forehead	20.1	7.1	3.9	9.1	7.5	**11.2**
	Eyes	36	21.8	**27.9**	19.3	22.1	27.5
	Nose	10.3	14.3	19.0	17.3	**9.4**	20.7
	Lips	4.6	7.3	9.1	9.7	**4.0**	9.0
	Hair	4.7	0.0	0.0	0.0	**3.4**	0.3
	Neck	3.3	1.0	0.4	**2.8**	0.7	0.4
	Left ear	2.9	0.8	0.0	0.5	**4.1**	1.3
	Right ear	0.9	2.3	0.2	**1.1**	2.3	**1.1**
	Chin	4.6	5.8	2.6	**5.5**	3.6	3.1
	Left cheek	3.8	6.6	1.3	**4.3**	8.0	6.2
	Right cheek	5.2	7.7	3.8	7.1	6.2	**5.3**
	Forehead	19.3	12.3	8.2	7.0	**16.1**	11.2
	Eyes	40.2	22.9	27.3	23.5	21.2	**31.0**
	Nose	8.0	15.6	24.5	14.5	**11.8**	15.5
	Lips	2.6	6.3	9.8	7.1	**5.2**	6.0
	Hair	5.8	0.2	0.0	1.0	**2.1**	0.2
	Neck	2.5	7.8	1.1	**2.8**	**2.2**	0.8
	Left ear	1.2	**1.2**	0.0	0.9	4.5	2.0
	Right ear	1.0	2.1	0.4	2.4	2.2	**1.2**
	Chin	2.0	8.1	3.5	8.1	5.0	**2.9**
	Left cheek	1.8	7.4	4.4	9.9	6.5	**4.3**
	Right cheek	2.4	5.6	**5.4**	9.3	6.8	8.3
	Forehead	16.5	8.0	7.2	**11.2**	**11.2**	8.8
	Eyes	37.9	20.1	**27.4**	19.0	21.1	22.9
	Nose	8.5	15.5	31.5	18.1	**12.5**	19.3
	Lips	4.2	9.7	12.4	10.7	**6.7**	8.3
	Hair	14.7	5.0	1.5	5.8	**15.2**	7.0
	Neck	2.8	5.1	1.4	1.8	5.2	**2.8**
	Chin	2.2	7.4	**3.6**	5.5	4.8	4.1
	Left cheek	2.3	7.8	**3.4**	4.6	8.5	7.7
	Right cheek	2.7	7.0	**5.3**	5.9	9.5	8.7
	Forehead	22.2	8.9	4.3	**10.0**	8.8	6.9

	Eyes	24.7	19.9	28.8	19.9	17.7	**23.7**
	Nose	9.0	17.7	20.6	16.4	**9.8**	14.0
	Lips	4.5	11.5	8.5	8.7	**6.7**	9.6
	Hair	7.4	0.3	0.1	0.8	**3.6**	2.2
	Neck	2.5	4.5	**2.1**	2.0	0.4	0.5
	Left ear	1.7	0.2	0.0	0.3	2.9	**0.6**
	Right ear	1.4	1.6	0.4	2.2	2.1	**1.2**
	Chin	2.0	10.4	6.5	9.5	**5.5**	7.8
	Left cheek	2.5	4.6	**1.8**	4.4	7.1	4.4
	Right cheek	2.7	6.7	**4.5**	5.6	5.6	7.0
	Forehead	26.6	12.2	8.8	**16.8**	13.8	11.0
	Eyes	31.0	20.6	**28.1**	21.3	17.8	24.5
	Nose	8.9	11.0	24.2	16.0	**10.7**	14.1
	Lips	4.3	8.4	7.0	9.7	4.9	**4.0**
	Hair	7.3	0.2	0.0	0.8	**4.6**	1.5
	Neck	2.7	**3.2**	0.7	**2.2**	1.2	0.0
	Left ear	2.7	1.0	0.1	1.3	3.9	**1.8**
	Right ear	1.4	2.5	0.8	**1.9**	3.2	2.1
	Chin	2.3	5.6	4.3	5.8	5.2	**2.7**
	Left cheek	3.5	7.4	**2.8**	6.7	8.6	7.1
	Right cheek	3.9	7.3	**6.4**	8.9	7.3	8.7
	Forehead	22.9	11.2	7.5	**15.2**	11.5	10.6
	Eyes	36.8	22.8	**24.8**	21.5	15.4	24.2
	Nose	10.1	15.1	22.8	18.1	**13.2**	16.8
	Lips	8.5	10.3	10.3	11.4	7.8	**8.6**
	Hair	4.0	0.2	0.1	0.9	**1.9**	0.2
	Neck	2.5	**2.7**	1.1	**2.7**	0.0	1.3
	Left ear	3.9	0.2	0.0	0.4	**3.1**	1.1
	Right ear	3.5	1.4	0.1	1.2	**2.0**	0.9
	Chin	4.1	9.3	**6.1**	8.4	7.7	6.8
	Left cheek	3.6	4.5	2.3	**2.9**	10.3	5.2
	Right cheek	4.1	5.7	3.4	**4.5**	6.5	7.5
	Forehead	15.5	10.3	6.0	**14.6**	12.0	12.3
	Eyes	28.4	20.3	25.1	23.1	22.7	**25.2**
	Nose	15.6	18.6	25.2	**16.3**	22.4	18.5
	Lips	7.5	2.3	13.1	10.4	11.0	**8.8**
	Hair	14.5	3.7	1.1	**7.7**	3.5	6.9
	Neck	3.7	**3.2**	0.9	0.7	1.8	1.8
	Chin	2.9	6.3	3.9	5.0	4.9	**2.8**
	Left cheek	2.3	3.8	**2.2**	6.3	5.3	5.5
	Right cheek	4.9	**5.3**	**5.3**	5.4	8.3	6.7
	Forehead	21.2	9.5	5.9	**10.4**	7.8	5.9

	Eyes	28.4	22.2	30.3	19.0	20.8	**27.3**
	Nose	15.6	**16.1**	26.5	19.2	10.4	13.2
	Lips	7.5	**8.7**	8.8	9.5	4.8	5.1
	Hair	4.9	7.1	2.0	**5.0**	15.3	12.7
	Neck	21.2	**3.6**	0.7	2.6	1.2	1.3
	Chin	3.7	6.4	**3.3**	4.2	2.3	2.0
	Left cheek	14.5	4.3	2.1	6.4	**7.8**	4.3
	Right cheek	2.3	**5.3**	6.4	5.7	8.1	7.3
	Forehead	2.9	8.4	**6.5**	11.9	11.0	9.8
	Eyes	27.3	21.7	**32.5**	21.4	16.4	21.4
	Nose	15.1	**17.0**	28.3	19.4	10.1	19.6
	Lips	5.4	10.0	**7.2**	9.5	7.6	8.0
	Hair	6.1	**9.1**	0.0	0.6	2.5	0.0
	Neck	4.0	2.7	1.9	1.1	**3.5**	1.6
	Left ear	2.2	**0.7**	0.0	0.6	4.1	**0.7**
	Right ear	2.4	1.6	0.5	1.4	**2.7**	1.5
	Chin	3.8	6.6	3.2	6.5	8.4	**3.8**
	Left cheek	6.1	**6.1**	2.5	3.9	7.7	4.8
	Right cheek	6.1	6.8	4.2	5.3	**6.2**	7.0
	Forehead	13.7	2.7	5.9	**12.3**	12.2	10.1

Observation Time in Relation to Character Gender

When comparing the average observation time for each area of interest (AOI) by character gender, slight differences were observed. The most notable difference was observed in the viewing of hair, where the average observation time for female characters' hair was approximately twice as long as that for male characters (Table 2).

Furthermore, in the case of male characters, more attention was allocated to the beard and cheek regions. Apart from the eye region – which consistently attracts the highest level of attention – these findings suggest that, compared to other facial areas, greater attention is directed toward the lower part of the face when observing male characters.

In contrast, for female characters, attention is more frequently focused on the upper part of the face.

When comparing the average observation times from GazePoint Analysis with the averaged predictions from AI tool, it was found that for male characters, the predictions show the highest correspondence in the *packaging* mode. For female characters, the *marketing material* mode shows a slight advantage in correspondence.

Table 2. Observation time based on the Virtual Humans' gender

AOI	Male Virtual Human (s)	Female Virtual Human (s)
Eyes	32.9	35.4
Nose	10.9	10.2
Lips	5.8	5.0
Hair	5.7	11.3
Neck	3.0	3.3
Left ear	2.4	1.6
Right ear	1.8	1.3
Chin	3.3	2.5
Left cheek	4.0	2.6
Right cheek	4.1	3.6
Forehead	18.7	19.2

Key Findings

Based on the conducted analyses, the following key findings can be highlighted:

- AI models effectively detect the primary regions of visual attention on virtual characters.
- A strong positive correlation exists between AI-predicted attention maps and results from traditional eye-tracking, indicating substantial agreement in measured visual attention patterns between the two methods.
- The largest differences appear in the outer areas of the stimuli.
- AI models become more accurate when observation times are longer.
- AI models usually show a less precise, more blurred pattern of attention than eye-tracking data.

These results show that AI tools can reliably capture overall patterns of visual attention, though they have some limits in precision.

Discussion

This study aimed to examine whether AI-based attention prediction tool can reliably replicate the visual attention behaviors found with eye-tracking when people observe virtual humans.

The results indicate a strong correspondence between the two methods, especially in spotting main areas of attention. This is consistent with previous research that points out the importance of these regions in face percep-

tion (Iskra & Tomc, 2016; Calvo & Nummenmaa, 2008). It confirms that AI models can capture key aspects of visual attention.

However, the differences found, especially in the outer facial areas and in how precisely attention is mapped, show the limits of AI-based methods. Eye-tracking records real user behavior as it happens, while AI models make predictions based on patterns from data. Because of this, AI models give a simpler and more spread-out picture of attention.

One key finding is that the match between the two methods improves with longer observation times. This suggests that AI models perform better at predicting steady phases of visual attention, but have trouble with the quick, early eye movements that often happen without conscious thought. This may be because early visual attention is more affected by personal and situational factors, which are hard for AI models to generalize.

In practice, these results indicate that AI tools can be a useful and effective alternative to eye-tracking in the early stages of design, when quick and affordable checks of user attention are needed. They are especially helpful in areas like UI/UX design, marketing, and virtual human development.

However, for studies that need a very precise and detailed analysis of behavior, eye-tracking is still essential. In this way, AI tools should be seen as a complement to, not a replacement for, eye-tracking, which corresponds to the results of earlier studies (Lavdas et al., 2021; Lengyel et al., 2021; Šola et al., 2024).

The results partly confirm the hypothesis: there is a strong match between AI and eye-tracking results, but there are clear limits in how precise and interpretable the AI results are.

Conclusion

The findings show a high level of agreement between the two methods, suggesting that AI tools can closely approximate overall patterns of visual attention. However, some limitations were observed, particularly in the precision of attention mapping, the analysis of peripheral areas of the stimuli, and the prediction of early observation phases.

For this reason, AI tools cannot fully replace eye-tracking, but they can be considered a useful complementary method for studying visual attention.

These results also point to the practical value of AI tools in the early stages of design, where quick and cost-effective evaluation of user attention is needed, especially in UI/UX design, marketing, and virtual character development.

Directions for future research include:

- The use of multiple AI tools and their combinations
- Analysis of dynamic stimuli (video and animation)
- Inclusion of additional metrics and biometric data
- Investigation of the influence of individual user characteristics on attention patterns

Further improvements in AI models, particularly through integration with real-world data, could help reduce existing differences and support their broader application.

Acknowledgement

This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-34/2026-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through project "Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad 2026" (No. 01-3609/1).

List of references

1. Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. Journal of Experimental Psychology: General, 137(3), 471–494. https://doi.org/10.1037/a0012771
2. Cornia, M., Baraldi, L., Serra, G., & Cucchiara, R. (2018). Predicting human eye fixations via an LSTM-based saliency attentive model. IEEE Transactions on Image Processing, 27(10), 5142–5154. https://doi.org/10.1109/TIP.2018.2851672
3. Duchowski, A. T. (2017). Eye tracking methodology: Theory and practice. Springer.
4. Golubović, G., Dedijer, S., Keresteš, N. M., Pál, M., Kerac, J., & Đurđević, S. (2026). On the influence of virtual characters' facial features on young adult user perception: A quantitative study. International Journal of Humanoid Robotics. Advance online publication. https://doi.org/10.1142/S0219843626500015
5. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. Oxford University Press.
6. Iskra, A., & Tomc, H. G. (2016). Eye-tracking analysis of face observing and face recognition. Journal of Graphic Engineering and Design, 7(1), 5–11. https://doi.org/10.24867/JGED-2016-1-005

7. Kümmerer, M., Theis, L., & Bethge, M. (2014). Deep gaze I: Boosting saliency prediction with feature maps trained on ImageNet. arXiv preprint. https://doi.org/10.48550/arXiv.1411.1045
8. Lavdas, A. A., Salingaros, N. A., & Sussman, A. (2021). Visual attention software: A new tool for understanding the "subliminal" experience of the built environment. Applied Sciences, 11(13), 6197. https://doi.org/10.3390/app11136197
9. Lengyel, G., Carlberg, K., Samad, M., & Jonker, T. R. (2021). Predicting visual attention using the hidden structure in eye-gaze dynamics. In CHI 2021 Eye Movements as an Interface to Cognitive State (EMICS) Workshop Proceedings. ACM.

FROM SOCIAL INTERACTION GRAPHS TO KNOWLEDGE GRAPHS: ENABLING GRAPHRAG FOR LLM-BASED ANALYSIS OF ONLINE CONVERSATIONS

Arber Ceni, Alda Kika
University of Tirana, Albania

Abstract

Large language models (LLMs) offer great support for the analysis of social media data; however, they often struggle to provide grounded and reliable responses when reasoning over social network structures. This limitation mainly results from their dependency on unstructured text which can lead to hallucinations or incomplete interpretations of complex interaction patterns. Retrieval-augmented generation (RAG) has been shown to improve grounding by including external knowledge; however, the existing approaches mainly use collections of documents and fail to exploit the relational and temporal structure already present in social media data. The aim of this research is to propose a framework for transforming social interaction graphs into knowledge graphs that can serve as context graphs for GraphRAG, enabling more grounded and interpretable LLM-based reasoning over online conversations.

The proposed framework starts from a social interaction network constructed from X (formerly Twitter) data, where nodes represent users and edges represent interactions between users such as mentions, replies, reposts (retweets), and quotes. This interaction graph is then transformed into a heterogeneous knowledge graph which has posts (tweets) at its center and explicitly models authorship, conversational structure, and temporal dynamics. Furthermore, this schema contains entities such as posts(tweets), users, hashtags, URLs, and conversations, and defines relations including authored, mentions, replies to, reposts (retweets), and quotes. Additional semantic enrichment is incorporated through entity extraction and linking, as well as metadata such as timestamps and community membership. The resulting knowledge graph can be used as a context graph within a GraphRAG pipeline, where relevant subgraphs are retrieved and provided to an LLM for answering questions.

This research highlights the potential of integrating social network analysis with knowledge graph construction to enhance LLM-based analysis of

social media data. By transforming interaction graphs into temporally enriched knowledge graphs, the proposed framework will enable structure-aware retrieval and reasoning, improving the quality and interpretability of generated responses.

Keywords: *GraphRAG, knowledge graph, large language models, social networks, X (formerly Twitter)*

Introduction

In recent years, large language models (LLMs) have experienced considerable advances in terms of architectural diversity, performance, usage and user adoption, leading to their integration in everyday tasks related to natural language processing and beyond (Brown et al., 2020). Social media platforms such as X (formerly Twitter), Facebook, Reddit produce large amounts of textual or visual content on which LLMs can reason and answer questions. However, users on these platforms do not only produce content, but also engage in different interactions between them including mentions, replies, reposts, quotes and shares. Such interactions form rich network structures which can be represented as social networks, where users are represented by nodes and the different interactions between them, form the edges of the network (Newman, 2003). These representations are often used in social network analysis but remain underutilized in LLM-based reasoning in contrast to unstructured textual content which continues to serve as the primary input of LLMs (Lewis et al., 2021).

Although they have strong language understanding capabilities, LLMs often struggle to provide reliable answers. When there is no clear path to the answer, LLMs often rely on assumptions, and this leads to incorrect or hallucinated responses (Brown et al., 2020). To alleviate hallucinations, retrieval-augmented generation (RAG) has been proposed. RAG works by incorporating external knowledge into the text generation process (Lewis et al., 2021). While RAG works well over unstructured text content, it fails to utilize the already present relational structure in social media interactions (Edge et al., 2025).

GraphRAG on the other hand, has demonstrated that modelling the knowledge as a graph, improves the grounding of LLMs by enabling multi-hop reasoning over nodes and edges of the graph (Lewis et al., 2021) (Peng et al., 2024). Current GraphRAG approaches are based on the textual data extracted from documents and the semantic relation between words to generate the knowledge graph (Peng et al., 2024). Social media data, on the oth-

er hand, naturally provide relational information which often is reduced into simple user-user interactions missing the underlying event context which can be used to reason upon (Ravi et al., 2024).

This paper explores the idea that social media interaction networks can be transformed into heterogenous knowledge graphs that serve as context graph for GraphRAG (Edge et al., 2025) (Peng et al., 2024). The proposed framework transforms a social network derived from X (formerly Twitter) posts, into a knowledge graph centered on tweets which models the conversational structure of the nodes in the original graph (users) and makes use of entities, hashtags and URLs. The constructed knowledge graph can support better reasoning by LLMs by preserving the structural and semantic dimensions of the original graph (Ravi et al., 2024).

Furthermore, a distinction between direct and indirect interactions such as mentions versus mentions through a retweet, reply-to or quote is made. This is done to maintain provenance and lower possible misleading interpretations. Temporal and community-level information also play an important role in capturing the dynamics of online conversations.

This work is positioned as a conceptual contribution, so our main focus is the design of the framework and the definition of evaluation criteria rather than empirical validation. A set of evaluation metrics for assessing grounding, interpretability, and reasoning capabilities in GraphRAG systems applied to social network data, and how these metrics can be used in future experimental studies is also discussed in this paper.

The contributions of this paper are threefold: (i) a framework for transforming social interaction graphs into knowledge graphs suitable for GraphRAG is introduced; (ii) the definition of a context graph representation that integrates structural, semantic, and temporal information; and (iii) the evaluation perspective for assessing LLM-based reasoning over social network data. This work contributes to the development of more reliable and explainable applications of LLMs in social computing.

Methodology

Social interaction graph construction

The initial dataset consists of several posts (tweets) related to COVID-19 vaccines downloaded from X (formerly Twitter) in a period between November 2021 and July 2022. For each term and for each month, a social graph was constructed based on the methodology described in (Ceni et al., 2023). Users who engaged in the conversation constitute the nodes in the graph and edges were built based on the different interactions those users

had. These interactions include mentions, reply-to, retweets and quotes allowing the capture of relational structure of the network. In Figure 1 is shown a social interaction graph created by analyzing 5 sample posts in X (formerly Twitter).

This representation is very effective for traditional social network analysis (Newman, 2003), but it fails to deliver information about the event itself. In this way, conversational structure, temporal dynamics and semantic content, which are important aspects of a social network, are lost. To overcome these limitations, a transformation of the original graph into a knowledge graph which focuses on posts instead of users and rebuilds all the underlying interactions is proposed.

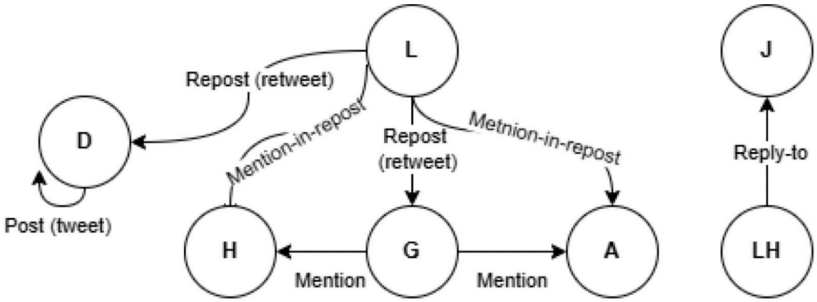

Fig. 1. Example of a social network created from X (formerly Twitter) data

Knowledge graph construction

The need to preserve structural and semantic dimensions, dictate the need to transform the social interaction graph into a knowledge graph (Peng et al., 2024). Edges which represent events around each individual post are now built from edges which represent user relationships. More specific, interactions between users are decomposed into one or more relations involving tweet nodes. If two users are connected by an edge in the original graph, those same two users are now connected through a set of intermediate edge and node entities that capture the content as well as the context of the interaction. This transformation represents the data more expressively because interactions are not anymore just edges between users, but structured events that can be used to reason upon.

The knowledge graph is now modelled as a heterogeneous graph consisting of different node and edge types.

The knowledge graph will contain these node types: (i) User – represents a user; (ii) Post – represents a post (tweet); (iii) Hashtag – represent a hashtag used in a post; (iv) URL and Domain – represent external references used in posts; (v) Entity – represents a real-world object extracted from post text; (vi) Conversation – grouping posts in a discussion thread; (vii) Community – represents a group of users identified through network structure

The knowledge graph will contain these edge types: (i) User → AUTHORED → Post; (ii) Post → MENTIONS → User; (iii) Post → REPLIES_TO → Post; (iv) Post → RETWEETS → Post; (v) Post → QUOTES → Post; (vi) Post → HAS_HASHTAG → Hashtag; (vii) Post → IN_CONVERSATION → Conversation; (viii) Post → HAS_MEDIA → Media; (ix) Post → MENTIONS_ENTITY → Entity; (x) Post → LINKS_TO → URL; (xi) URL → IN_DOMAIN → Domain; (xii) User → MEMBER_OF → Community

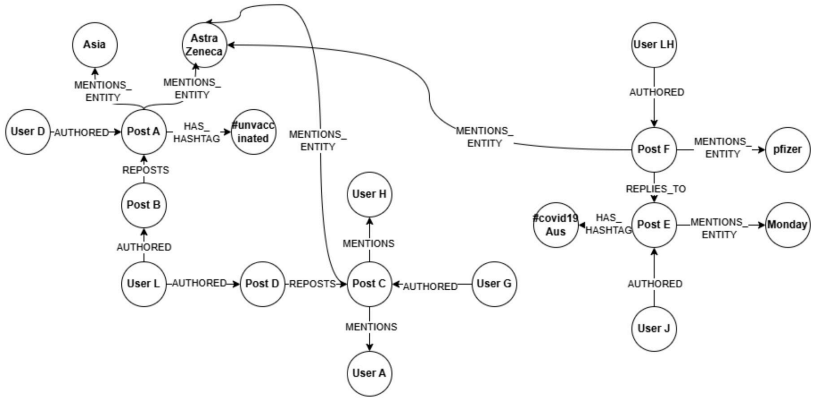

Fig. 2. The social network transformed into a knowledge graph following our proposed framework

Figure 2 shows the knowledge graph obtained by transforming the interaction graph of Figure 1 using the proposed framework.

This is a power framework to model different types of user interactions associated with their semantic. A particular aspect of this framework is the ability to distinguish between direct and indirect interactions. Direct interactions are considered interactions such as reply-to, mention, repost (retweet) and quote. Indirect interactions are considered mentions achieved through repost (retweet) or quote.

Instead of using a single relation type to represent all interactions and using attributes to distinguish them, the knowledge graph contains distinct relations (MENTIONS, REPLIES_TO, RETWEETS, QUOTES) thus preserving the semantic of the original interaction. Indirect interactions are stored in the structure of the graph itself and can be derived through multi-hop traversal thus always maintaining the origin of the interaction. The importance of this distinction is important because interaction through an intermediate step is not as important as a direct interaction. Users usually repost or quote without the primary intention of mentioning other users. We argue that by preserving this distinction, the graph will support better and more accurate reasoning.

Besides preserving the different users' interactions, the graph expressiveness is enriched by including additional semantic and contextual data generated from the post content and other metadata like: (i) Named entities – extracted from text using entity recognition techniques; (ii) Hashtags – representing different topics; (iii) URLS and domains – representing external references; (iv) Temporal information – timestamps when the post or the user profile was created; (v) Community information – derived from the graph's structure by using community detection algorithms.

The above elements allow the knowledge graph to not only capture interactions between users, but also *what* is being discussed, *when* interactions occur, and *how* the information travels across the network. So, this knowledge graph provides a richer context for downstream reasoning tasks (Ravi et al., 2024).

Context Graph for GraphRAG

One of the main differences between GraphRAG and classic RAG is that GraphRAG needs a context graph instead of context documents (Edge et al., 2025). The generated knowledge graph will serve to further generate context graphs for GraphRAG. When the user inputs a query, a subgraph is retrieved from the knowledge graph based on structural and semantic criteria which include user interactions, entities and the conversation context. The subgraph will serve as a structured context input for the language model, thus enabling it to generate responses based on relationships and evidence on the graph. By taking advantage of graph-based retrieval, this approach enables multi-hop reasoning on users, interactions and content and at the same time preserves interpretability given that nodes; and edges' connections are traceable.

Analysis

Analytical capabilities of the proposed framework

The proposed framework which transforms a social interaction graph into a knowledge graph enables us to perform richer analysis over social media content. The traditional user-user interaction network models relationships as aggregated edges while the transformed knowledge graph models interactions as event-level entities centered around posts. This way both structural relationships and their context is preserved.

As a result, more types of analytical queries are supported. For example, questions such as *"Which users interact most frequently"* can be answered using the structural aspect of the graph. At the same time, more complex queries like *"What topics are discussed between two communities over time?"* require the use of structural, semantic and temporal information, which are present in the entities of the knowledge graph. Answering these questions is not always possible and straightforward with the traditional graph representation.

Expected improvements in LLM-based reasoning

The use of the proposed knowledge graph as a context graph in the GraphRAG framework is expected to improve the quality of responses generated by LLMs in several ways. First, by limiting the model to operate on retrieved subgraphs, it will reduce the reliance of the LLM on assumptions, thus providing a better grounding (Lewis et al., 2021). Second, the model will be able to infer relationships that are not directly present in the graph or observable from pieces of text. This belief is based on the fact that the graph structure enables multi-hop reasoning across users, posts and other entities present in the graph (Edge et al., 2025). For example, the following query *"How are users A and B connected?"* may require traversing multiple nodes (conversation, mutual interaction, common hashtags/URLs). In a text-based RAG, such relationships may be difficult to reconstruct because they are not always present in the text entities.

Role of interaction semantics

In the proposed framework, the explicit modelling of interaction semantics like mentions, replies, reposts, and quotes, allow the knowledge graph to preserve the intent and context of user actions. This is an important aspect in social network analysis because different interaction types convey different meanings. Another important aspect of the design of the proposed framework is the separation between direct and indirect interactions. For

example, a direct user mention may indicate intentional engagement, while a mention through a repost or quote may not reflect direct intentional communication. By keeping this distinction, the framework allows for more accurate analysis of user relationships by reducing the risk of misleading interpretations.

Temporal and community-level reasoning

User interactions in social media are dynamic, and they evolve over time and across communities, so the inclusion of temporal information and community structure is of the essence, and it allows the framework to capture these dynamics (Li et al., 2025). This way, questions about the evolution of the conversation can be asked, such as *"How has the interaction between two users changed over time?"* or *"Which communities are driving a particular discussion?"*. The addition of temporal information allows the model to distinguish early and late interactions, thus supporting causal and sequential reasoning. The presence of community-level information helps the identification of different interactions patterns within and across groups.

Comparison with traditional approaches

Traditional social network representation such as user-user interactions, lack the ability to capture content, context and temporal dynamics. Text-based approaches, on the other hand, focus more on content but ignore relational structure. Our proposed framework sits in between and bridges this gap by integrating structural, semantic and temporal dimensions. This way it supports both network-level analysis and content-aware reasoning, making it a good candidate for integration with LLM-based systems.

Evaluation perspective

For the purpose of evaluating the proposed framework, the focus is on four key dimensions: (i) grounding, (ii) reasoning capability, (iii) structural awareness and (iv) interpretability. The evaluation will be performed as a comparative experimental setup involving three approaches: (a) vanilla LLM – the model generates responses without access to external context, (b) text-based RAG – textual documents (posts/tweets) are retrieved and provided as context to the model (Lewis et al., 2021) and (c) GraphRAG – a query-specific subgraph is retrieved from the knowledge graph and provided as context to the model (Edge et al., 2025). The same dataset and query set will be used throughout the evaluation process.

Grounding – tries to measure how well the generated responses are supported by evidence in the retrieved subgraph and how well the model can

avoid hallucinations (Peng et al., 2024). Two measures can be used to quantify grounding: (i) evidence-supported answer rate – the proportion of responses that can be attributed to nodes and edges in the retrieved subgraph, (ii) faithfulness score – the degree of consistency of the generated response compared to the supporting graph structure

Multi-hop reasoning – the ability of the model to identify indirect connections between users through shared conversations or common entities. Two metrics can be dsitinguished: (i) multi-hop accuracy – measures if responses to these queries are correct, (ii) path validity – determines if the identified relationships are valid paths in the graph.

Temporal reasoning – evaluates the dynamicity of the model by calculating the following measures: (i) temporal consistency – measures if responses follow the ordering and timing of interactions, (ii) temporal reasoning accuracy – evaluates if the model can answer time related questions regarding interactions (Li et al., 2025).

Structural awareness – tries to determine if the model correctly interprets the different interaction types present in the graph. The following metrics can be used to measure structural awareness: (i) relation correctness – does the models identify and use the correct interaction type? (ii) edge-type differentiation – is the model able to distinguish between direct and indirect interactions?

Interpretability – determines whether responses given by the model can be linked to the graph structure. (i) Traceability – if the response can be linked to specific entities on the graph.

Hallucination rate – can be regarded as the proportion of responses that are not supported by the retrieved subgraph. It provides an estimate of the model's reliability (Peng et al., 2024).

The following query categories are defined to evaluate the proposed framework:

- Relational queries – *"Which users interact more frequently?"*
- Multi-hop queries – *"How are users A and B connected?"*
- Semantic queries – *"What topics are discussed within a community?"*
- Temporal queries – *"How has the interaction between two users evolved over time?"*

Limitations and open challenges

The proposed approach shows great potential as a conceptual framework; however, it presents several challenges that need to be addressed when implementing it in future work. The construction of the knowledge graph relies on carefully extracting entities, interactions and metadata and

often may be affected by noise and ambiguity. Second, the size and complexity of the graph may introduce challenges for retrieval and reasoning. It is known that the effectiveness of GraphRAG depends on the quality of the retrieved subgraph (Peng et al., 2024). Retrieving the optimal subgraph while maintaining completeness is not a trivial task. The evaluation of the reasoning over graph-structured data needs carefully designed benchmarks and metrics, aspects which are outlined in the following section.

Conclusions

Large language models have shown limitations when applied to social network data, especially in producing grounded and reliable responses. Although retrieval-augmented generation has improved the integration of external knowledge, it focuses mainly on text and does not take advantage of the inherent relation structure found in social media interactions.

In this paper, a framework for transforming social interactions graphs derived from social media into a heterogeneous knowledge graph which will then serve as a context graph within GraphRAG is proposed. A post-centered construction of the knowledge graph which explicitly models interaction semantics, temporal dynamics and community structure, thus preserving structural and semantic dimensions of social media data is further discussed. This will contribute to a richer analysis and more grounded and interpretable responses by LLMs.

Our primary contribution in this position paper is the conceptual design of the framework as well as the definition of an evaluation perspective focused on the graph-based reasoning over social media data. To this purpose, five evaluation dimensions, including grounding, multi-hop reasoning, temporal reasoning, structural awareness and interpretability are outlined.

The implementation of the proposed framework and the execution of the described experiments is the focus of our future work. This includes the creation of benchmark datasets, the design of retrieval strategies for constructing the context graph, and the comparison of GraphRAG approaches against traditional LLM and text-based RAG solutions. This line of research has the potential to significantly improve the reliability and explainability of LLM-based analysis in social computing.

List of references

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are*

Few-Shot Learners (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165
2. Ceni, A., Kika, A., & Çollaku, D. X. (2023). *A Social Network Analysis of COVID-19 Vaccines Tweets.* 1–12.
3. Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, R. O., & Larson, J. (2025). *From Local to Global: A Graph RAG Approach to Query-Focused Summarization* (arXiv:2404.16130). arXiv. https://doi.org/10.48550/arXiv.2404.16130
4. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks* (arXiv:2005.11401). arXiv. https://doi.org/10.48550/arXiv.2005.11401
5. Li, D., Niu, Y., Ai, Y., Zou, X., Qi, B., & Liu, J. (2025). T-GRAG: A Dynamic GraphRAG Framework for Resolving Temporal Conflicts and Redundancy in Knowledge Retrieval. *Proceedings of the 33rd ACM International Conference on Multimedia,* 11880–11889. https://doi.org/10.1145/3746027.3755628
6. Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167–256. https://doi.org/10.1137/S003614450342480
7. Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., Zhang, Y., & Tang, S. (2024). *Graph Retrieval-Augmented Generation: A Survey* (arXiv:2408.08921). arXiv. https://doi.org/10.48550/arXiv.2408.08921
8. Ravi, R., Ginde, G., & Rokne, J. (2024). *PRAGyan—Connecting the Dots in Tweets* (arXiv:2407.13909). arXiv. https://doi.org/10.48550/arXiv.2407.13909

A HYBRID MACHIN ELEARNING AND DEEP LEARNING SYSTEM FOR PHISHING EMAIL DETECTION IN WEBMAIL PLATFORMS

Elda Xhumari, Rivalda Bedini
University of Tirana, Albania

Abstract

Phishing emails continue to represent a major cybersecurity threat, exploiting social engineering techniques and deceptive language to compromise user credentials, financial data, and organizational systems. As phishing campaigns become increasingly sophisticated and automated, traditional rule-based detection methods struggle to identify new attack patterns. In response to this challenge, this research proposes a hybrid phishing email detection framework that combines classical machine learning and deep learning techniques for integration within webmail platforms. The objective of the study is not only to improve phishing detection accuracy but also to demonstrate the operational feasibility of deploying intelligent detection systems directly in a real-world email environment.

The proposed framework is trained on a multi-source dataset consisting of more than11,000 labeled emails balanced between legitimate and phishing messages. The dataset combines legitimate emails from the Enron Email Corpus with phishing samples obtained from public phishing repositories and AI-generated phishing emails. This diverse dataset allows the model to capture both traditional phishing patterns and emerging AI-generated attack styles. To enhance classification performance, the system integrates textual representations extracted from email subject and body with engineered structural indicators such as suspicious domain patterns, URL entropy, keyword flags, subdomain counts, and other metadata-based features.

Two modeling pipelines were implemented and evaluated. The first employs classical machine learning algorithms, including Logistic Regression, Random Forest, and Support Vector Machine, trained using TF-IDF textual features combined with engineered attributes. The second pipeline utilizes deep learning architectures, specifically Bidirectional Long Short-Term Memory (Bi-LSTM) networks and the DistilBERT transformer model, to capture contextual language patterns and semantic relationships present in phishing messages. DistilBERT was selected due to its balance between strong predictive capability and relatively low computational cost, enabling near real-time email analysis.

Experimental results demonstrate strong classification performance across all evaluated models using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The best-performing model was integrated into the Roundcube webmail platform to enable real-time phishing detection. Incoming emails are automatically analyzed and suspicious messages can be redirected to a dedicated phishing folder, demonstrating the practical applicability of hybrid AI-based detection systems for strengthening operational email security.

Keywords: *phishing detection, machine learning, deep learning, email security, hybrid classification.*

Introduction

Email phishing represents one of the most critical and escalating threats in modern cybersecurity. Through social engineering and psychological manipulation, attackers impersonate trusted entities, such as financial institutions or business associates, to deceive victims into disclosing sensitive credentials, clicking malicious links, or executing actions that result in financial or data loss. Phishing attacks exploit users trust and remain one of the most commonly used techniques in cybercrime, targeting both individuals and organizations (A. S. & A. H., 2023). The increasing prevalence of these attacks and their potential impact on organizations and individuals underline the need to develop effective detection and prevention methods, thus reducing the risk and costs derived from these incidents (Anti-Phishing Working Group, 2024). The scale of this threat continues to grow: phishing attacks reached 989,123 reported incidents in the fourth quarter of 2024 alone, representing the highest quarterly figure recorded that year, up from 932,923 in the third quarter and 877,536 in the second quarter, with a monthly average of approximately 330,000 attacks (Anti-Phishing Working Group, 2024).

Despite the existence of spam filters and rule-based defenses, such attacks continue to bypass conventional detection mechanisms. Traditional detection techniques often rely on predefined rules or signature-based approaches, which struggle to identify newly emerging phishing strategies or AI-generated messages (A. S. & A. H., 2023). As phishing campaigns evolve and become more sophisticated, more advanced approaches are required to detect malicious emails effectively.

The aim of this research is to develop and evaluate an intelligent phishing email detection system using hybrid machine learning (ML) and deep learning (DL) approaches, and to demonstrate its practical applicability

through deployment in a real webmail environment. The object of the study is email-based phishing detection using textual and structural email features. To achieve this aim, the following objectives were defined: (1) constructing a comprehensive multi-source labeled dataset of over 11,000 emails from sources including Enron, Public Phishing Email Corpus, OpenPhish, and AI-generated samples; (2) implementing classical ML classifiers, namely Logistic Regression, Random Forest, and Support Vector Machine, with hybrid TF-IDF and engineered feature inputs; (3) developing DL architectures including a Bidirectional LSTM and a fine-tuned DistilBERT transformer; (4) comparatively evaluating all models using accuracy, precision, recall, F1-score, and ROC-AUC metrics; and (5) integrating the best-performing model into the Roundcube webmail platform for real-time phishing detection.

The research employs an empirical methodology combining quantitative model evaluation with system integration testing, using scikit-learn for ML models and TensorFlow/Keras for DL architectures, applied across both real-world and synthetically generated email datasets.

Methodology and equipment

This study follows an empirical research design aimed at developing and evaluating a hybrid phishing email detection system. The central hypothesis is that combining textual features with engineered structural and URL-based attributes produces superior classification performance compared to single-modality approaches.

The dataset was constructed from multiple sources to ensure diversity and coverage. Legitimate emails were drawn from the Enron Email Corpus, a well-established collection of real corporate communications (cs.cmu.edu/~enron). Phishing samples were sourced from a public phishing email corpus containing real attack emails (academictorrents.com), and supplemented with approximately 2,800 AI-generated phishing emails (huggingface.co/datasets/kxm1k4m1/generate_phishing_email_final). The final dataset contains 11,388 labeled email samples, with 6,002 legitimate (52.7%) and 5,386 phishing (47.3%), achieving a near-balanced class distribution suitable for model training without resampling.

Data preprocessing involved text normalization including lowercasing, removal of non-ASCII characters, stripping of reply chains, signature blocks, and legal disclaimers, as well as filtering of empty or invalid entries. Hash-based deduplication was applied to eliminate redundant samples across sources.

Feature engineering produced 28 numeric and boolean attributes per email, capturing URL complexity indicators such as entropy, path depth,

domain count, and maximum URL length, alongside content-based signals including subject length, special character count, and keyword flags. Textual features for classical ML models were represented using TF-IDF vectorization applied separately to the subject and body fields, while deep learning models used tokenized sequences with padding for uniform input length.

Five models were developed and evaluated: Logistic Regression, Random Forest, and Support Vector Machine using scikit-learn, alongside a Bidirectional LSTM and a fine-tuned DistilBERT transformer implemented in TensorFlow/Keras and HuggingFace Transformers respectively. All models used a hybrid input combining textual and engineered numeric features. ML models were validated using k-fold cross-validation, while DL models used train/validation/test splits with binary cross-entropy loss, dropout regularization, and early stopping.

The complete workflow was implemented in Python 3.10, with Pandas and NumPy for data processing, Matplotlib and Seaborn for visualization, and the Roundcube webmail platform for real-world system integration and deployment testing.

Presentation of research results (Analysis)

System architecture and modeling approach

The phishing detection system is built on a hybrid input architecture that combines textual features extracted from email content with engineered numerical and boolean indicators. The overall framework operates across three stages: preprocessing, model training, and deployment. Two parallel modeling pipelines were developed and evaluated: a classical machine learning pipeline comprising Logistic Regression, Random Forest, and Support Vector Machine, and a deep learning pipeline comprising a Bidirectional LSTM and a fine-tuned DistilBERT transformer. Both pipelines share a common preprocessing phase and are trained on the same dataset to ensure a fair comparison.

For the ML pipeline, email subject and body text were vectorized using TF-IDF with unigram and bigram ranges, and combined with standardized engineered features through scikit-learn's ColumnTransformer. For the DL pipeline, the Bi-LSTM used a Keras tokenizer with a vocabulary of 20,000 words and sequences padded to 256 tokens, while DistilBERT used the HuggingFace WordPiece tokenizer with identical sequence length. In both DL architectures, a dual-branch design was adopted: a text branch producing contextual embeddings and a numerical branch processing engineered features, with outputs concatenated at a fusion layer before final classification. Training used binary

cross-entropy loss, Adam optimizer, dropout regularization, early stopping, and class weighting to address label imbalance.

Model performance results

All five models were evaluated on a held-out test set of 1,703 emails using accuracy, precision, recall, F1-score, and ROC-AUC. Results are summarized in Table 1.

Table 1. Final Performance Metrics for Each Model (Test Set)

Model	Accuracy	Precisio	Recall	F1-	ROC-AUC
Logistic Regression	99.47%	99.26%	99.63%	99.44%	0.9999
Random	99.59%	99.26%	99.88%	99.57%	1.0000
SVM	99.30%	99.13%	99.38%	99.25%	0.9999
BiLSTM	99.76%	99.66%	99.88%	99.77%	1.0000
DistilBERT	99.94%	99.87%	100.00%	99.94%	0.9999

All models surpassed 99% across every metric, confirming the effectiveness of the hybrid feature design. DistilBERT achieved the highest overall performance, producing zero false positives and missing only a single phishing email across the entire test set, an error rate of just 0.06%. BiLSTM followed closely with an error rate of 0.23%, misclassifying only 4 emails in total. Among classical ML models, Random Forest was the strongest performer with 99.59% accuracy and only 7 total misclassifications, while SVM produced the most errors at 12 misclassifications, though still achieving 99.30% accuracy. ROC-AUC values of 0.9999 or above across all models indicate near-perfect class separation regardless of threshold.

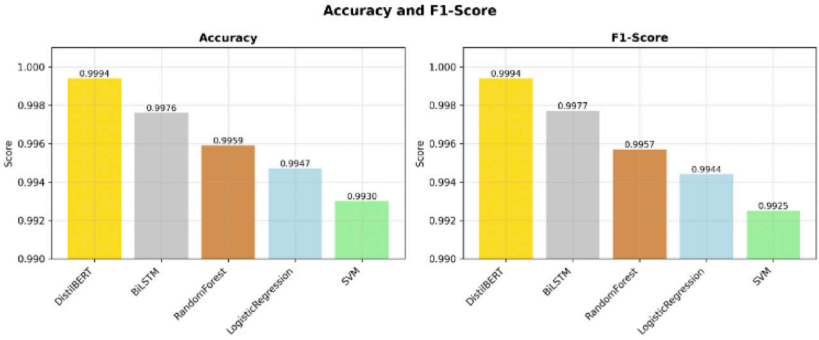

Fig. 1. Comparison of Model Accuracy and F1-Score Across All Models

ML vs. DL comparison

Grouping models by approach reveals a clear advantage for deep learning in minimizing false positives. The ML group produced 19 total false positives across all three classifiers, compared to only 1 for the DL group. Mean accuracy for DL models was 99.85% compared to 99.45% for ML models, and mean F1-score was 99.86% versus 99.42%, as shown in Table 2.

Table 2. *Comparison of ML vs. DL Models*

Group	Mean Acc.	Mean Prec.	Mean Recall	Mean F1	FP	FN	Best Model
Machine Learning	99.45 %	99.22 %	99.63 %	99.42 %	19	9	Random Forest
Deep Learning	99.85 %	99.76 %	99.94 %	99.86 %	1	4	DistilBER T

Random Forest closely approached DL-level recall, missing only one phishing email, and offers the additional advantage of feature interpretability, where URL entropy, number of domains, and keyword flags emerged as the most influential predictors. ML models are significantly lighter in computational requirements, making them more suitable for resource-constrained deployment environments. Deep learning is the preferred choice when accuracy and reliability are paramount, while machine learning remains attractive in constrained environments due to speed, interpretability, and simplicity (Uddin et al., 2024).

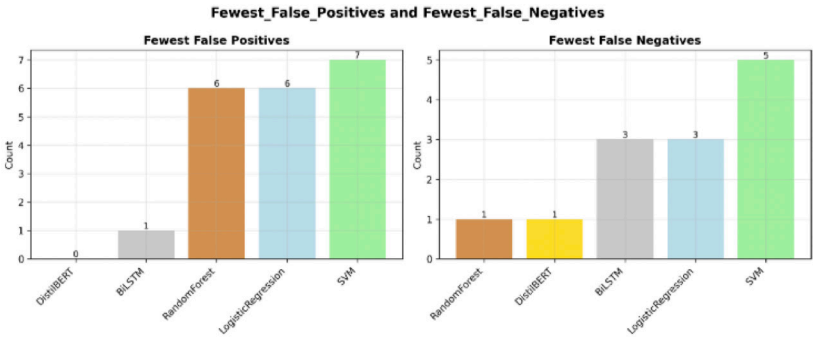

Fig. 2. *False Positive and False Negative Counts Across All Models*

Production testing

The best-performing model was integrated into the Roundcube webmail platform for real-time phishing detection. Two deployment configurations were tested: a DistilBERT-only branch and a hybrid RF + DistilBERT branch implementing a two-stage pipeline where Random Forest pre-filters emails before passing borderline cases to DistilBERT for deeper semantic analysis. During live testing with a balanced mix of real and simulated phishing emails, both branches successfully detected all phishing attempts. The DistilBERT-only branch occasionally produced isolated false positives on legitimate emails containing multiple embedded links. The hybrid branch demonstrated greater consistency, maintaining zero false positives across repeated test sessions while preserving accurate phishing detection, confirming its robustness and suitability for long-term production deployment.

Error Analysis

Despite near-perfect performance, a small number of misclassifications occurred across all models, summarized in Table 3. Phishing emails that were missed typically employed subtle social engineering without overt URL manipulation, closely mimicking legitimate communication styles. False positives were primarily legitimate emails containing multiple hyperlinks or urgent transactional language, which inadvertently triggered phishing indicators such as high URL entropy or keyword flags.

Table 3. Breakdown of Prediction Outcomes (TN, FP, FN, TP) by Model

Model	TN	FP	FN	TP	Total	Error Rate
Logistic Regression	893	6	3	801	1703	0.53%
Random Forest	893	6	1	803	1703	0.41%
SVM	892	7	5	799	1703	0.70%
BiLSTM	898	1	3	801	1703	0.23%
DistilBERT	898	0	1	804	1703	0.06%

These findings suggest that future improvements should incorporate URL reputation checking, adversarial training samples, and larger transformer architectures to address edge cases where semantic subtlety outweighs structural signals.

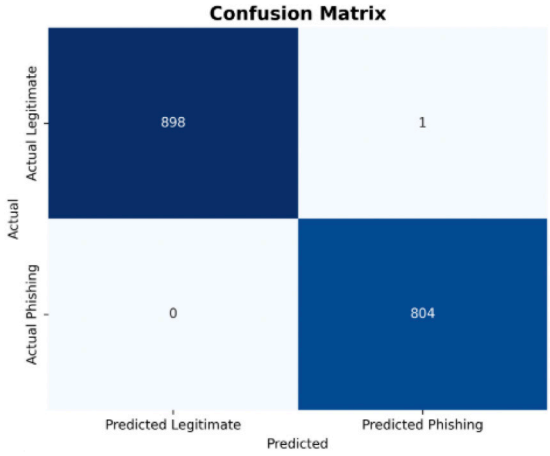

Fig. 3. DistilBERT Confusion Matrix (Test Set)

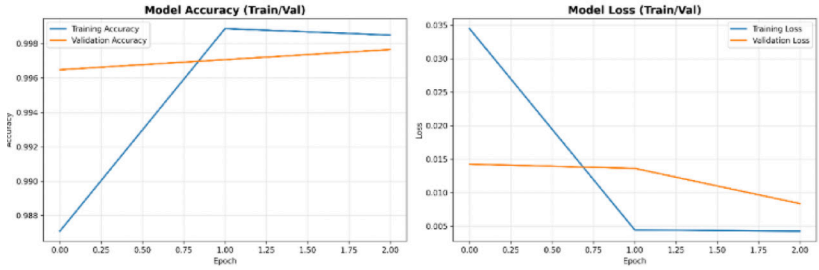

Fig. 4. DistilBERT Training and Validation Accuracy and Loss Curves

Conclusions

This study developed and evaluated an intelligent phishing email detection system combining classical machine learning and deep learning approaches within a unified hybrid architecture. The research confirmed that integrating engineered structural and URL-based features with contextual textual embeddings produces superior detection performance compared to single-modality approaches.

All five models achieved over 99% accuracy on the test set. DistilBERT delivered the strongest results with 99.94% accuracy, zero false positives, and only one missed phishing email, confirming the advantage of transformer-based architectures in capturing semantic nuance (Songailaite et

al., 2023). BiLSTM followed closely with an error rate of 0.23%, while Random Forest proved the most competitive classical model, approaching deep learning performance while offering interpretability and lower computational cost.

Real-world deployment in the Roundcube webmail platform demonstrated the system's practical feasibility. The hybrid RF + DistilBERT production configuration maintained zero false positives across repeated live testing sessions, confirming its stability and suitability for operational email security environments.

For practical deployment, the following recommendations are proposed. First, a risk-based classification approach, categorizing emails into severity levels rather than binary outcomes, enables more nuanced organizational response. Second, detection thresholds should be calibrated conservatively to minimize false positives and preserve user trust. Third, continuous retraining with recent phishing campaigns is essential given the rapid evolution of attack strategies, particularly AI-generated phishing. Fourth, the system should complement rather than replace user awareness, with clear explanations of flagged emails reinforcing human judgment alongside automated detection.

Future work should address multilingual detection to extend applicability beyond English datasets, adversarial robustness against evasion tactics such as obfuscated URLs and modified phishing text, and the integration of larger transformer architectures balanced with distillation techniques for efficient deployment. Incorporating multimodal signals including images, sender metadata, and logos, alongside integration with enterprise security infrastructures such as SIEM platforms, represents a promising direction for comprehensive next-generation phishing defense.

List of references

1. A. S., & A. H. (2023). Phishing emails detection model using deep learning.
2. Anti-Phishing Working Group. (2024). Phishing activity trends report: 4th quarter 2024. Anti-Phishing Working Group.
3. Songailaitė, M., Kankevičiūtė, E., Zhyhun, B., & Mandravickaitė, J. (2023). BERT-based models for phishing detection. In Proceedings of the 28th International Conference on Information Society and University Studies (IVUS 2023).
4. Uddin, M. A., Islam, M. M., Hossain, M. S., & others. (2024). An explainable transformer-based model for phishing email detection: A large language model approach.

INNOVATIONS IN PUBLISHING, PRINTING
AND MULTIMEDIA TECHNOLOGIES

INOVACIJOS LEIDYBOS, POLIGRAFIJOS
IR MULTIMEDIJOS TECHNOLOGIJOSE

Vol 2 No 1 (2026)

ISSN 2538-8622

Published by *Kauno kolegija Higher Education Institution*
Layout *Virginijus Valčiukas*