Raising user awareness about the consequences of online photo sharing

Hugo Schindler, Adrian Popescu, Khoa Nguyen, Jérôme Deshayes-Chossart

Raising User Awareness about the Consequences of Online Photo Sharing

Hugo Schindler
hugo-schindler@orange.fr
Université Paris-Saclay, CEA, List
F-91120, Palaiseau, France

Adrian Popescu
adrian.popescu@cea.fr
Université Paris-Saclay, CEA, List
F-91120, Palaiseau, France

Van-Khoa Nguyen
van-khoa.nguyen@etu.unige.ch
University of Geneva
Geneva, Switzerland

Jérôme Deshayes-Chossart
jerome.deshayes-chossart@cea.fr
Université Paris-Saclay, CEA, List
F-91120, Palaiseau, France

ABSTRACT

Online social networks use AI techniques to automatically infer profiles from users' shared data. However, these inferences and their effects remain, to a large extent, opaque to the users themselves. We propose a method which raises user awareness about the potential use of their profiles in impactful situations, such as searching for a job or an accommodation. These situations illustrate usage contexts that users might not have anticipated when deciding to share their data. User photographic profiles are described by automatic object detections in profile photos, and associated object ratings in situations. Human ratings of the profiles per situation are also available for training. These data are represented as graph structures which are fed into graph neural networks in order to learn how to automatically rate them. An adaptation of the learning procedure per situation is proposed since the same profile is likely to be interpreted differently, depending on the context. Automatic profile ratings are compared to one another in order to inform individual users of their standing with respect to others. Our method is evaluated on a public dataset, and consistently outperforms competitive baselines. An ablation study gives insights about the role of its main components.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Security and privacy** → **Privacy protections**.

KEYWORDS

images, graph neural networks, user awareness, privacy

ACM Reference Format:
Hugo Schindler, Adrian Popescu, Van-Khoa Nguyen, and Jérôme Deshayes-Chossart. 2023. Raising User Awareness about the Consequences of Online Photo Sharing. In *International Conference on Multimedia Retrieval (ICMR '23), June 12–15, 2023, Thessaloniki, Greece.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3591106.3592290

ICMR '23, June 12–15, 2023, Thessaloniki, Greece

ACM ISBN 979-8-4007-0178-8/23/06…$15.00
https://doi.org/10.1145/3591106.3592290

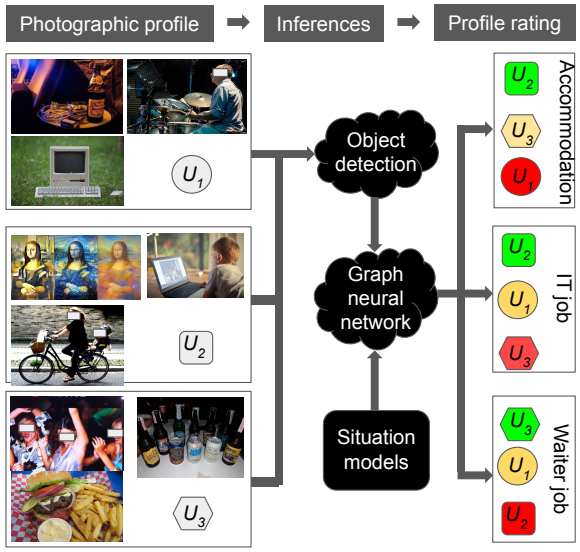


Figure 1: Potential consequences of personal data sharing for 3 users in 3 real-life situations. Users' photos include information about their interests and lifestyle which can be inferred and aggregated to compute the appeal of a profile in context. The aggregation is done with a graph neural net trained to rate profiles. The network is trained with object detections and ratings per situation and uses human ratings of training profiles as ground truth. Modeling several situations is useful because the same shared information can be interpreted differently. For instance, the two top row photos of U_1 and U_3 might be assessed negatively when searching an accommodation, but positively for a waiter job. Top row photos of U_2 are positive for IT job, but their photos with young kids might be problematic if searching for a waiter job. A ranking of profile ratings is used to raise awareness about the consequences of data sharing.

1 INTRODUCTION

Users are entitled to know how the data they share online can be leveraged by online social networks (OSNs) and third parties. Despite its importance for an informed online participation, the

proposal of efficient user feedback remains difficult due to a combination of usability and technical challenges. Efforts to increase transparency and trust, such as Facebook's Privacy Checkup [13], were recently made under public and regulatory pressure but their effectiveness is strongly debated [5, 51]. The availability of on-device deep learning models [38, 44] enables the creation of AI-assisted tools which provide user feedback before sharing. This is important insofar as feedback is most efficient before sharing, when data are still on the user's device [41].

Users should be able to understand the potential effects of their data-sharing practices. This topic was explored during the ImageCLEF 2021 Aware Task [20] which introduced a dataset which includes photographic user profiles, associating object detections and human ratings of these profiles in four real life situations. The modeled situations include search for: an accommodation, a bank loan, an IT job and a waiter job. Several situations are needed since the effects of data sharing vary depending on the context in which they are used, as illustrated in Figure 1. The objective of the Aware task was to automatically rate a set of photographic user profiles per situation in order to reproduce human ratings. The focus was on lightweight algorithms which can be run on users' devices. Inferences are thus done before the actual sharing, and this is important insofar as OSNs gain control of the data once they are shared.

We summarize the proposed method for rating photographic profiles in Figure 1. We note that the rating of these profiles varies in the three illustrated situations. For instance, the analysis of their shared photos is likely to place U_2 at the top of the user profile rankings for situations accommodation and IT job search because their shared photos are more appealing than those of U_1 and U_3 in the two situations. Inversely, U_2 will be low ranked for waiter job search because their photos are either unrelated to this situation or can be judged negatively. Note that the ratings of the profiles encode social biases of the people who perform the ratings [20, 31]. The automatic rating process will inherit these biases, but this is not considered problematic in our approach, because it is meant to simulate real-life situations, including biases. The rating process is based on Graph Neural Networks (GNNs) [40], whose ability to learn from entities and their relations in a flexible way [4] makes them particularly suited to our use case. We explore different ways to structure data as graphs and three GNN models to adapt the automatic rating of profiles to each situation. We evaluate the proposed model on the publicly available ImageCLEF Aware dataset [20], which contains a total of 500 profiles. Results show a consistent improvement compared to a newly proposed baseline and to an existing method. Further experiments show that more performance gains are obtainable by using ensembling techniques.

2 RELATED WORK

The main objective here is to raise users' awareness regarding their online activity, and more precisely about personal data sharing practices. Our work is inspired by advances in relevant disciplines, including legal research, behavioral economics, and psychology. Legal research is relevant insofar as it examines the practical implications of the legal framework, such EU's General Data Protection Regulation, on the design of tools which provide feedback to users [5, 28, 51]. Behavioral economics motivates our approach since they point out real-life effects of OSN participation. These studies are often focused on the influence of sensitive information, such as gender, origin, or religion, on hiring practices [1, 29, 45]. Beyond legal safeguards against their discriminatory use, the influence of strong social signals is rather well grasped by the users. While following a similar global objective, we infer weaker signals by an automatic analysis of personal data because their effects are not well understood. Recent studies in psychology [30, 42] advocate for the need for transparency in order to facilitate an informed participation in OSNs and, ultimately, to build a relation of trust between services and their users in the long run. Our contribution improves transparency by providing insights both about automatic inferences and their effects.

Photos constitute a large part of the data shared on OSNs. A wealth of information can be inferred from their content and used against users [48]. Existing works usually classify user images as private/public. An early solution [50] used shallow classifiers on top of a bag-of-visual-words representations. More powerful deep representations [25] were then used to improve accuracy. A recent review of deep models for privacy prediction [47] concludes that pretrained ResNet extractors work best. Adapted deep representations were equally proposed to improve predictions [18]. The authors of [41] introduced a personalization component which accounts for personal variations of privacy perception. Personalized predictions with multi-layered semantic graphs to encode both common and personal views of privacy were proposed in [22]. These works are interesting insofar as they raise user awareness. However, they differ from our approach because they stop short of explaining the effects of sharing. In our study, we give a contextualized view of sharing consequences by showing that the same profile is perceived differently depending on the situation.

We build on the ImageCLEF 2021 Aware shared evaluation task [20], which proposes a dataset for raising user awareness. A set of 500 photographic profiles sampled from the YFCC dataset [46], itself collected from Flickr, were manually rated in 4 real-life situations. Detections of visual objects for each profile, along with crowdsourced object ratings in each situation, are provided. A solution to this task which uses regression and profile descriptor compression was introduced in [31]. Another solution based on random forest regressors was introduced recently [32]. An ensemble of estimators is deployed over a representation of profiles which combines object detection confidence and situation scores of objects. The approach is fine tuned and best results are obtained with an ensemble of 650 estimators. While interesting, these solutions either do not leverage recent advances in deep learning. Another recent approach [26] tested the use of a dense neural network for the task However, the reported results are poor.

Graph neural networks (GNNs) [40] are suited to model the task. They learn from data which can be structured as graphs which encode entities and their relations. GNNs are able to learn complex information about entities, their associated relations and their compositionality rules. This differentiates them from other types of deep learning models [4]. They store information as embeddings and rely on graph network operations, based for instance on Message Passing Neural Networks, to update these embeddings during learning [17]. Graph Convolution Networks [24] use convolutions adapted for graphs to update node features based on the properties

	ACC	BANK	IT	WAIT
ACC	1.00	**0.77**	0.54	0.22
BANK		1.00	0.54	0.19
IT			1.00	**0.10**
WAIT				1.00

Table 1: Pearson correlation coefficients (ρ) between the ground truth ratings of the four modeled situations in the ImageCLEF 2021 Aware dataset [20]

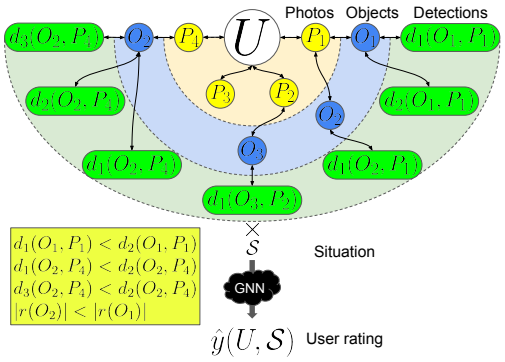


Figure 2: Proposed graph-based inference process. A visual profile is structured as a graph. Here, the user has four photos. Photo P_1 has three detections: two object O_1 and of O_2. P_2 has a single detection of O_3. P_3 does not have any detection. P_4 has three detections of O_2. The features of this graph are then enriched based on the situation objects scores. A GNN is finally used to get $\hat{y}(U, \mathcal{S})$, the predicted user profile rating in the situation.

of the neighbors. Two or three layers are typically used in [24] to share information locally and to not disturb the global pooling. Our learning objective is defined as a graph regression which is done after the graph network update. It requires an aggregation step which is done via a pooling operation. This aggregation step is challenging since the type and order of pooling will influence the result [49]. A preliminary analysis indicates that GNN performance comparison is challenging [10, 12] and leads us to retain three types of aggregation for experiments. The aggregation step which follows the convolutional layer is classically done using a mean pooling operation. Self-attention graphs (SAG) [27] introduce an attention mechanism to improve the pooling step. Graph Isomorphism Network (GIN) [49] obtain robust performance in classification by improving the neighbor aggregation step [49].

3 PROPOSED CONTRIBUTION

3.1 Motivation

We propose GUAR, a method for GNN-based User Awareness Raising. We motivate it and present its components. Users share their personal data on OSNs in an initial context which they choose and control. These data are then aggregated into profiles which can be used in secondary contexts which are were initially anticipated. A known example is the sharing of geolocated photos of holidays which, in some cases, led to the users' houses being burglarized [16]. Albeit rare, this example illustrates a serious consequence of sharing an information which seems innocuous. More systematic usage of knowledge inferred from users' online contributions is possible in situations such as those modeled in the ImageCLEF 2021 Aware dataset [20]. Even though such usages can be legally problematic, they do occur in practice [5], The Aware dataset includes four situations to convey the contextualized interpretations of shared data. We illustrate these different interpretations in Table 1 by presenting the correlations between the ground truth ratings of the four situations. The Pearson correlation varies between 0.1 and 0.77 for the IT-WAIT and ACC-BANK pairs, respectively. WAIT is also decorrelated from ACC and BANK since the corresponding Pearson correlation coefficient ρ values are 0.22 and 0.19, respectively. The differences between situations call for their adapted processing in order to optimize performance. Consequently, we propose a flexible learning framework both in terms of GNN architecture, learning process parametrization and data processing.

3.2 Notations

We define the main notations used in the formalization of the content of the ImageCLEF 2021 Aware dataset [20]. A user profile U_i is defined as a set of photos: $\mathcal{U}_i = \{P_1^i, P_2^i, ..., P_n^i\}$. The visual detection is composed of a set of objects: $\mathcal{D} = \{O_1, O_2, ..., O_w\}$. A situation is represented as a set of crowdsourced ratings of the detectable objects: $\mathcal{S} = \{r(O_1), r(O_2), ..., r(O_w)\}$. Note that the ratings of an object will vary across situations [31]. The set of visual profiles included in the dataset is: $\mathcal{U} = \{U_1, U_2, ..., U_x\}$. The ground truth rating of user U_i in $\mathcal{S}$, $y(U_i, \mathcal{S})$, is also crowdsourced. The predicted rating is $\hat{y}(U_i, \mathcal{S})$. $d_k(O_l, P_j^i)$ is the confidence score of k^{th} detection of object O_l in the j^{th} image of user U_i. k is necessary to account for multiple detections of an object in a photo.

3.3 GUAR description

The structure of the Aware dataset can be modeled using a graph which links users, their images and detected objects, as illustrated in Figure 2. A star-like shape was retained because it suits the structure of the data. A user U is the core of the graph and is linked to the photos P_j included in the photographic profile, which are themselves connected to visual objects O_l. Finally, O_l nodes are connected to detections represented by $d_k(O_l, P_j)$. The number of objects detected is variable. Some photos include no detectable object, while other contain several of them. An advantage of graph-based modeling is its flexibility. Any type of node can be removed when searching for optimal representations of profiles in a situation. For instance, if nodes P_j are removed, U will be directly linked to object nodes O_l. If the U node is removed, the graph will include multiple disconnected subgraphs, each one with P_j as root node. Different graph structures, which include all or only a part of the nodes presented in Figure 2, are tested during the learning process. After selecting the graph backbone, it is necessary to preprocess

the object detections and encode the features in the graph. These steps are presented hereafter.

Detection preprocessing. A threshold can be applied to retain detections whose confidence $d_k(O_l, P_j^i)$ is such that a good balance between kept true positives and removed false positives is found. Since the detection range varies for different objects O_l, it is interesting to define adapted thresholds. We follow [31] and optimize thresholds $\eta(O_l)$ using a filtering-based attribute selection mechanism [35]. This mechanism maximizes the Pearson correlation between automatic and manual user ratings over the training set when object detectors are activated individually. We also tested other thresholds, such as the average or median predictions in the training subset, but they gave lower performance.

Annotators use winner-take-all strategies [21] in order to rate photographic profiles. Since it is impossible to know at which level they are deployed, we implement three strategies to restructure the graph from Figure 2 and illustrate them in Figure 3. Max pooling is applied to different graph components to model winner-take-all. We first hypothesize that, given several detections of O_l, only the most confident one is likely to matter when the image is analyzed during profile annotation. We implement this selection strategy in the MP_1 variant from Figure 3(a). When MP_1 is used, each retained detection is represented using a one-hot vector in the feature encoding described in Subsection 3.3. On top of MP_1, we can assume that only the object which has the largest situation score $r(O_l)$ (in absolute value) among those detected in the image matters. This strategy, named MP_2, is illustrated in Figure 3 (b). Finally, MP_3 also builds on MP_1 but max-pooled detections are merged in a multi-hot vector. These pooling strategies are compared to MP_0, the variant without pooling illustrated in Figure 2.

Feature encoding. Multi-hot encoding is used in order to link parent nodes with their children nodes from Figure 2 whenever the edges between them are active. The activation is defined by the preprocessing steps described in Subsection 3.3. The object detection nodes are represented using either one-hot encoding for MP_0, MP_1 and MP_2 pooling strategies or multi-hot encoding for MP_3. The object ratings $r(O_l)$ can be used in order to contextualize the detection nodes $d_k(O_l, P_j)$ of the graph representation for each modeled situation $\mathcal{S}$. Then, $r(O_l) \times d_k(O_l, P_j)$ is used instead of $d_k(O_l, P_j)$ in the coding of the detection nodes.

GNN models. We use graph neural networks to learn how to predict user ratings based on the user graph structures described above. We evaluated three GNN models since their a priori comparison for a new task is difficult [10, 12]. The first model, named MEAN, includes a number of layers of convolutional layers, a mean pool layer followed by a multilayer perceptron (MLP). The second model, named SAG, also uses convolutional layers, but they are followed by three SAGPool layers [27] which implement a self-attention mechanism. The aggregation is done by concatenating max and mean pooling layers, and by an MLP. The third model, named GIN, is a graph isomorphism model [49] composed of dense layers, followed by a sum pool and by an MLP. Based on the GNN selected, we introduce three models: $\text{GUAR}_{\text{MEAN}}$, GUAR_{SAG}, and GUAR_{GIN}.

The models are trained using hyperparameters related to the graph architecture, the user graph preprocessing, and the characteristics of the training process. Except for the loss which is defined below, these hyperparameters are listed in Subsection 3.3 and discussed in more details in the supplementary material. Given a subset of training profiles from $\mathcal{U}$, the objective of the learning process is to maximize the correlation between automatic automatically obtained user rating and those from the ground truth. Since the Pearson correlation coefficient is used to measure performance in the ImageCLEF 2021 Aware task [20], we also use it in the training process. This coefficient is defined as:

$$\rho_{Y,\hat{Y}} = \frac{\mathbb{E}[(Y - \mu_Y)(\hat{Y} - \mu_{\hat{Y}})]}{\sigma_Y \sigma_{\hat{Y}}} \tag{1}$$

with: ρ - the Pearson correlation coefficient, Y - the set of ground truth ratings of user profiles, $\hat{Y}$ - the set of predicted ratings of user profiles, $\mathbb{E}$ - the expectation and μ, σ - the mean and standard deviation, respectively.

We want to maximize the correlation between Y and $\hat{Y}$, therefore we define the loss as $\mathcal{L}(Y, \hat{Y}) = -\rho_{Y,\hat{Y}}$.

4 EXPERIMENTS

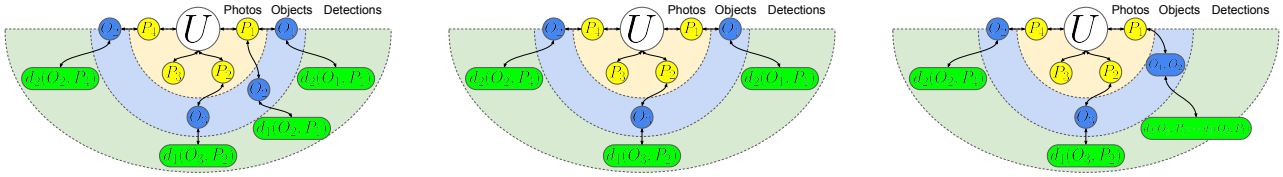
Evaluation is done with a publicly available dataset [20]. Our solution is compared to methods from [31] but also to a new competitive baseline introduced here. An analysis of different method components and parameters is provided in order to understand their respective contributions.

4.1 Dataset and metrics

The ImageCLEF 2021 Aware dataset [20] includes 500 photographic user profiles sampled from the YFCC dataset [46], with 100 images per profile. Manual ratings of profiles per situation are obtained through crowdsourcing, and are used as ground truth. Profiles are characterized by two main elements. First, object detections with Faster-RCNN [37] was done using a dataset of 269 objects. The total number of detections in the dataset is 67,938. Second, object ratings per situation were obtained through crowdsourcing and are available for use. The dataset is split in 360, 40, 100 profiles for train, validation, and test. The evaluation objective is to produce an automatic ranking of profiles per situation which is as close as possible to the manual ground truth. The Pearson correlation coefficient is used to measure the correlation of human and automatic profile ratings for the test set. Correlations in individual situations are aggregated to provide a global score.

4.2 Baselines

To our knowledge, the only public solution which tackles the task was introduced in [31]. Two methods from [31] are tested here, along with a new baseline AUTO_{SKL}. BASE_{η} - represents photographic profiles as 269-dimensional vectors which combine object detections and ratings in each situation. Test profiles are ranked by summing individual object detections weighted by situation-specific object ratings. LERVUP^{fr} - compresses full profile vectors using object rating positivity, negativity, and average detection confidence. A focal rating component gives more weight to objects



(a) MP_1 pooling (b) MP_2 pooling (c) MP_3 pooling

Figure 3: Max pooling strategies applied to the full graph and the relations defined in Figure 2. (a) MP_1 keeps only the most confident detection per object. (b) MP_2 keeps only the max confidence detection of the object with the highest absolute value rating $r(O_l)$ per image. (c) MP_3 merges most confident object detections per image. This creates multi-object detections.

Model	ACC	BANK	IT	WAIT	Average
BASE$_\eta$ [31]	0.31	0.34	0.41	0.46	0.38
LERVUP^{fr} [31]	0.43	0.36	0.54	0.59	0.48
AUTO$_{\text{SKL}}$	0.27	**0.44**	0.56	0.68	0.49
GUAR$_{\text{SAG}}$	0.38	0.27	0.44	0.62	0.43
GUAR$_{\text{GIN}}$	0.44	0.20	0.54	0.59	0.44
GUAR$_{\text{MEAN}}$	**0.54**	0.36	**0.57**	**0.71**	**0.55**

Table 2: Pearson correlation coefficients ρ on the test split for the ImageCLEF 2021 Aware dataset [20] with the baselines and the proposed method. GUAR performance is reported with the GNN models described in Subsection 3.3. ρ interpretation is done using the ranges from [9]: weak for $[0.1, 0.3]$, moderate for $[0.3, 0.5]$, strong for $[0.5, 1]$. *Best results for individual methods are in bold.*

with have high ratings, which are likely to be the most influential. Profiles are automatically rated using a random forest model for regression with the compressed representations. AUTO$_{\text{SKL}}$ - uses auto-sklearn [15] to search suitable models in Scikit-learn [34]. Similar to $BASE_\eta$, profiles are represented as 269-dimensional vectors. AUTO$_{\text{SKL}}$ is interesting because it finds an optimal classical learning model.

4.3 Implementation

We used TensorFlow Geometric [19] since it enables an easy deployment of GNNs. The main hyperparameters of GUAR(Subsection 3.3 that were explored are: (1) the network architecture: type of deep graph network, number of layers, size of layers, type of nodes included in the architecture, directionality of the edges; (2) the data preprocessing: max pooling type, feature encoding type, adaptive thresholds for object detections, removal of outlier training profiles; (3) the learning process: optimizer, learning rate range policy, batch size. An efficient exploration of these parameters is performed using a Bayesian method from Optuna [2]. More details are provided in the supp. material.

4.4 Main experiment

The results from Table 2 indicate that the performance of GUAR$_{\text{MEAN}}$ is interesting because the correlation between manual and automatic profile rating is strong for ACC, IT and WAIT and moderate for BANK, following the intervals given in [9]. Equally important, GUAR$_{\text{MEAN}}$ provides a consistent performance improvement compared to the baselines. Globally, gains of 6 and 7 Pearson correlation coefficient ρ points are obtained compared to AUTO$_{\text{SKL}}$ and LERVUP^{fr}, respectively. The results vary significantly between situations. WAIT is the easiest situation and BANK the most difficult one for GUAR$_{\text{MEAN}}$. The authors of [31] explained this variation, also observed for LERVUP^{fr}, by the fact that WAIT is better represented in the object detection dataset than the other situation. This explanation is related to the automatic inferences which are associated to a user's photos. A further explanation might be related to the quality of the ground truth data associated to each situation. Their judgments about the appeal of each profile in a situation is based on an aggregation of weak signals from profile photos. These signals are probably more interpretable in some situations than in others. GUAR$_{\text{MEAN}}$ has the best results among the three types of GNN models tested. GUAR$_{\text{SAG}}$ and GUAR$_{\text{GIN}}$ lag consistently behind, with average performance drops of 12 and 11 points with respect to GUAR$_{\text{MEAN}}$. This is interesting insofar as SAG and GIN aggregation components were proposed to improve over MEAN but are not efficient here. This finding might be explained by the size of the available training set. A simpler GNN architecture seems better suited here but further comparisons should be performed if the dataset is extended.

AUTO$_{\text{SKL}}$ has the best behavior among the baselines, followed by LERVUP^{fr} and BASE$_\eta$. This result shows that the feature engineering component proposed in [31] is not necessary if an appropriate learning method is selected. AUTO$_{\text{SKL}}$ even provides the best performance for BANK, with an 8-points gain compared to GUAR$_{\text{MEAN}}$. The information available for BANK is probably insufficient for an efficient encoding using a deep graph network.

4.5 Ablation of training users and objects

The number of rated user profiles and the number of object detections are the most important features of the training set. We remove 25% and 50% of users or objects randomly to analyze their influence. The results from Table 3 confirm that both features are important. The performance is lower when a larger number of profiles and objects are removed. Performance drops reach 2 and 6 points with 25% and 50% of users removed and 5 and 12 points with 25% and 50% of objects removed. The performance drop is higher for object ablation compared to profile ablation, particularly when half of each is removed. Globally, the results from Table 3 indicate that

	ACC	BANK	IT	WAIT	Average
$GUAR_{MEAN}$	0.54	0.36	0.57	0.71	0.55
User ablation 25%	0.52 ±0.05	0.36 ±0.03	0.54 ±0.04	0.70 ±0.00	0.53 ±0.02
User ablation 50%	0.43 ±0.06	0.32 ±0.03	0.50 ±0.04	0.69 ±0.03	0.49 ±0.02
Object ablation 25%	0.48 ±0.06	0.34 ±0.06	0.49 ±0.05	0.70 ±0.02	0.50 ±0.03
Object ablation 50%	0.38 ±0.06	0.27 ±0.10	0.39 ±0.08	0.68 ±0.03	0.43 ±0.06

Table 3: Performance after the ablation of 25% and 50% of user profiles and visual objects from the training subset of ImageCLEF 2021 Aware dataset [20]. Users and objects are ablated randomly five times. The average Pearson correlation coefficient ρ, and the standard deviations are reported.

	ACC	BANK	IT	WAIT	Average
$GUAR_{MEAN}$	0.54	0.36	0.57	0.71	0.55
MP_0 pooling	0.48	0.27	0.57	0.74	0.51
MP_1 pooling	0.54	0.33	0.57	0.73	0.54
MP_2 pooling	0.52	0.36	0.55	0.71	0.54
MP_3 pooling	0.52	0.36	0.57	0.67	0.53

Table 4: Performance obtained by fixing the type of max pooling used in all four situations. Max pooling notations are introduced in Subsection 3.3. MP_0 means no pooling.

performance should be improved if the dataset was richer. The performance reduction is lowest for WAIT among the four situations, a situation which is well represented. The degradation is particularly important for IT and ACC when 50% of objects are ablated. These situations are fairly well represented in the full dataset, but this representation is strongly degraded with fewer objects. Inversely, these situations are probably the ones whose performance could be improved the most is new objects which are relevant for them would be available.

4.6 Effect of max pooling

Max pooling was introduced to simulate winner-take-all strategies used by human raters [21]. In Table 4, we present results without max pooling and with the three max pooling schemes introduced in Subsection 3.3. In each case, the pooling strategy is fixed for all four situations. This is to be compared to a flexible choice of the optimal pooling method per situation whose results were reported in Table 2. All three max pooling flavors improve results compared to MP_0, which does not perform pooling. The best performance is obtained by MP_1 and MP_2, which apply pooling at object and object plus image levels, respectively. Their scores are just 1 point below of that of $GUAR_{MEAN}$, which selects the best pooling type for each situation. The pooling operation has variable influence in the four modeled situations. IT is the least affected, while differences are strongest for BANK. Note that $GUAR_{MEAN}$ has lower performance than MP_0 and MP_1 for WAIT since best results for validation and test subsets are not fully aligned. Both subsets should be richer in order to further improve results.

Ensemble method	ACC	BANK	IT	WAIT	Average
$AUTO_{SKL,\ ENS}$	0.47	**0.47**	0.57	**0.72**	0.56
SSNCSE [32]	0.44	0.45	0.54	0.7	0.53
$GUAR_{MEAN,\ KERAS}$	**0.56**	0.35	0.60	0.69	0.55
$GUAR_{MEAN,\ SKL}$	**0.56**	0.36	0.62	0.67	0.55
$GUAR_{SAG,\ ENS}$	0.54	0.38	**0.67**	0.71	0.58
$GUAR_{GIN,\ ENS}$	0.44	0.39	0.60	0.59	0.51
$GUAR_{MEAN,\ ENS}$	**0.56**	0.42	0.66	0.70	**0.59**

Table 5: Pearson correlation coefficients ρ on the test split for the ImageCLEF 2021 Aware dataset [20] with ensembling methods. ρ interpretation is done using the ranges from [9]: weak for $[0.1, 0.3]$, moderate for $[0.3, 0.5]$, strong for $[0.5, 1]$. Best results for individual methods are in bold.

4.7 Effect of ensembling models

We tested several ensembling techniques and describe them briefly hereafter: (1)$AUTO_{SKL,\ ENS}$ uses auto-sklearn [15, 34] to combine multiple models available in Scikit-learn [34], whereas $AUTO_{SKL}$ only selects a single model. (2) SSNCSE is a method which was recently introduced in [32], and uses an ensemble of regressors to for the 2022 edition of the Aware task. For fairness, we report results obtained with it for the 2021 version of the dataset, which is used in this work. (2) $GUAR_{MEAN,\ KERAS}$ is based on the rating predicted by the top-validation $GUAR_{MEAN}$ models. The ratings are combined by a neural network optimized by auto-Keras [8, 23]. This approach is motivated by the recent reports in [43] regarding the usefulness of deep nets as post-processors in ensembling. (3) $GUAR_{MEAN,\ SKL}$ is based on the rating predicted by the top-validation $GUAR_{MEAN}$ models. The ratings are ensembled by an ensemble of traditional machine learning models optimized by auto-sklearn [15, 34]. This is a "classical" learning counterpart of $GUAR_{MEAN,\ KERAS}$. (4) The models $GUAR_{SAG,\ ENS}$, $GUAR_{GIN,\ ENS}$, and $GUAR_{MEAN,\ ENS}$ are based on $GUAR_{SAG}$, $GUAR_{GIN}$, and $GUAR_{MEAN}$, respectively. They are combined using the averaged rescaled rating method tested in the main experiments.

The performance of the different ensembling techniques is summarized in Table 5. $AUTO_{SKL,\ ENS}$ and SSNCSE [32] are two ensembling methods which do not use deep learning and are similar in spirit since they aggregate the outputs of linear classifiers. $AUTO_{SKL,\ ENS}$ has better performance since it explores and combined a more diversified range of classifiers. $GUAR_{MEAN,\ KERAS}$ and $GUAR_{MEAN,\ SKL}$, the methods which exploit learning-based post-processors do not work in our task. This might be because the data provided to them is not sufficient for learning in an effective way. $GUAR_{MEAN,\ ENS}$, introduced in Subsection 4.2, provides a 4-points improvement over the use of a single $GUAR_{MEAN}$ model. It is notably beneficial for BANK and IT, which gain 6 and 9 points over

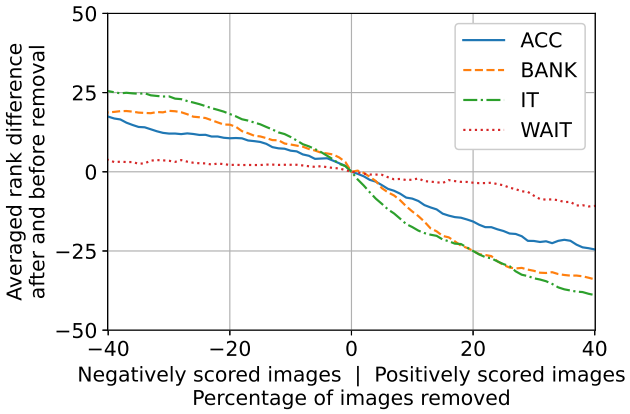


Figure 4: Image removal effect on the profile rating evaluated by its relative position in a ranked set of a 100 reference profiles. We use a leave-one-out procedure in the test set to obtain target and reference profiles. We remove a variable proportion of negatively and positively scored images per profile. The ranking difference after and before removal is averaged over the test set and presented per situation. Profile ranks are improved and are degraded if negatively scored images (left) and if positively ranked images (right) are removed. This behavior confirms the capacity of the proposed method to provide feedback to a user about the effect of their images in different situations.

single models. We note that WAIT does not benefit from ensembling, and a small improvement is obtained for ACC. Ensembling is useful in practice if the additional computations needed to run multiple models can be afforded. For instance, the optimal configuration for BANK and IT is obtained by combining 5 and 6 individual models, respectively. The smaller these numbers are, the more interesting the ensembling is. More details about the number of models used by ensembling are given in the supplementary material.

4.8 Ablation of profile images

GUARcan be used to enable an AI assistant to inform a user about the effects of photo sharing in real-life situations. We highlight these effects by removing positively or negatively scores images from the profile. The resulting ablated profile can be compared to a set of reference profiles, which remain unchanged, to assess how its ranking evolves. The test subset, except for the target profile, is used as reference in this experiment. The target profile ranking should become better and worse, respectively, when negatively and positively scored images are removed. The rank change will then give the user an understanding of the effects of removed images. Individual image scores are computed using an aggregation of object situation ratings and of detection confidence from [31]:

$$sc(P) = \sum_{l=1}^{w} r(O_l) \sum_{k=1}^{L} d_k(O_l, P) \quad (2)$$

with: P - the scored image, w - the cardinality of the detection dataset $\mathcal{D}$, L - the number of detections of O_l in P.

The results obtained by removing between 0% and 40% of negatively and positively scored images of each profile are presented in Figure 4. The effect of the removal, quantified via the user profile rank change, is important since the removal of 40 negatively scored images (left of Figure 4) results in an average positive change of the rating of up to 25 places. Inversely, the removal of positively rated images (right of Figure 4) results in an average negative change of the profile rating of up to 33 places.

Rank changes are not symmetrical when removing the same amount positively and negatively scored images. This result is due to the fact that object ratings in situations are not symmetrically distributed. The profile rank changes vary among situations. WAIT ranks are least affected by image removal because, as explained in Subsection 4.5. This happens because WAIT has a lot of pertinent visual objects in a lot of images. Also, the ratings of these objects are generally closer to neutral than in the other situations. The expected effect of each individual image is consequently smaller. Thus, a larger amount of profile images can be removed with relatively low effect on the ranking. The largest variations are obtained for BANK and IT. This happens because there are fewer relevant images which shape profile ratings in these cases, but with stronger ratings associated to each of them. The obtained results confirm that $\text{GUAR}_{\text{MEAN}}$ captures the potential effects of sharing negatively and positively scored images. It can thus help the user make informed decisions about which images should be shared and which should be kept private.

5 DISCUSSION

GUARtackles a task which has both technical and societal implications. It is thus important to discuss limitations and risks, along with mitigation strategies when available. A first limitation is related to the size of the ImageCLEF Aware dataset [20]. The ablation experiments from Section 4 indicate that a larger dataset would be beneficial, both in terms of number of user profiles and of detectable objects. We hope that this dataset will be expanded and the result will be made available to the community. However, the ablation also shows that the obtained results are stable, and the conclusions drawn from experiments are valid. More generally, the availability of larger and more diversified datasets which can be used to provide feedback about data sharing is desirable. However, their creation is not straightforward due to legal, ethical and technical challenges.

A second limitation is related to the content of the dataset, which is focused photo-related predictions. The proposed algorithm was instantiated for this type of data but its adaptation to other types of inputs is simple due to the flexible nature of graph neural networks. Note also that focus here is on soft signals extracted from shared data, while existing work presented in Section 2 focuses on stronger signals, such as demographic characteristics of users. The two types of signals should naturally be integrated in order to provide thorough feedback to users.

A third limitation is that of technical and social biases which are encoded in the dataset. This problem was discussed in detail by the authors of [31]. They point out that a series of measures were adopted to reduce biases, notably technical, whenever possible. However, it is difficult not to encode human biases, such as the subjectivity of human ratings of photographic profiles and of

visual objects. Such biases are inherent to any annotation process which appeals to the participants' opinion and will be inherent to any downstream task which use human ratings. One common mitigation measure, also implemented for the Aware dataset, is to obtain multiple ratings per item and thus transform personal biases into social ones. We acknowledge the presence of biases and believe that explanations about them should be included in the feedback provided to users.

We aim to improve the algorithmic core of the awareness raising pipeline which is supported through the ImageCLEF Aware task [20]. It is interesting to discuss its implications regarding explainability, a key feature for the proposal of trusted AI-driven products and services. The authors of [3] analyze different types of explanations (textual, visual, local, by example, by simplification, by feature relevance). The implementation of these types of explanations is often directed toward an understanding of AI models work. Such strategies for explainability can be implemented for GNN models [14, 36] and would be useful for ML experts and practitioners. However, they might be less adapted for non-expert users despite their tendency to overvalue the technically oriented explanations [11]. Explanations for non-experts, such as OSN users, are more effective if they are embedded in a practical context and directly linked to their goals. The use of several situations to show the consequences of sharing is related to explainable AI by projection [39] since the effect of input changes (i.e. profile images) can be accessed via the user ranking per situation. A potential implementation of this type of explanations discussed in Subsection 4.8, with the removal of negatively/positively scored images. This process enables the users to make more informed decisions about which data should be shared and which should stay private. The use of situated effects is in contrast with existing photo privacy assistance methods [33, 41, 47, 50] which do not contextualize their recommendations.

A potentially important risk is the misuse of our contribution by malicious third parties. This type of risk is well identified for a large spectrum of AI techniques which have societal implications [6]. However, algorithms that are probably more sophisticated than ours are already deployed by OSNs and associated third parties, with little attention given to transparency [7]. One can argue that, somewhat ironically, the misuse predates the virtuous use of technologies from the users' perspective. Also important, the means afforded by service providers are disproportionately high compared to those that we, as researchers, and final user have. OSNs should ideally increase transparency by providing AI-assisted tools which enable users to understand how their data are processed and aggregated. Since this does not seem a priority for them, it is important to propose such tools independently. Our work can be useful in practice if integrated in tools such as https://ydsyo.app, the app prototype supported via the ImageCLEF Aware task. The method was developed with such integration in mind and thus focused on inference methods which can be used at the edge. The code will be made publicly available.

6 CONCLUSION

We have introduced a new GNN-based method which raises user awareness about the consequences of personal data sharing in impactful situations. Our results indicate that it is possible to automatically rate users' photographic profiles in an effective way. These ratings can then be aggregated into ratings using a community of reference in order to provide feedback to users about how well they stand in the crowd. The proposed model leverages graph flexibility to represent user profiles and an adapted data structuring process to regress the corresponding ratings to improve results compared to competitive baselines. The experiments show that the use of GNNs results in a consequent improvement of performance compared to existing methods.

The promising results reported here encourage us to pursue our work. We will explore other types of deep networks, such as transformers, in order to further improve results. The focus was on images because the dataset provided for the task includes only this type of data. It would be interesting to study how other types of shared data (texts, videos, etc.) influence the perception of user profiles.

Acknowledgements. This work was supported by the European Commission under European Horizon 2020 Programme, grant number 951911 - AI4Media. This work was supported by the Fondation MAIF. It was made possible by the use of the FactoryIA supercomputer, financially supported by the Ile-de-France Regional Council.

REFERENCES

[1] Alessandro Acquisti and Christina Fong. 2020. An Experiment in Hiring Discrimination via Online Social Networks. *Management Science* 66, 3 (2020), 1005–1024. https://doi.org/10.1287/mnsc.2018.3269

[2] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, August 4-8, 2019*, Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis (Eds.). ACM, Anchorage, AK, USA, 2623–2631. https://doi.org/10.1145/3292500.3330701

[3] Alejandro Barredo Arrieta, Natalia Díaz Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* 58 (2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[4] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinícius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Çaglar Gülçehre, H. Francis Song, Andrew J. Ballard, Justin Gilmer, George E. Dahl, Ashish Vaswani, Kelsey R. Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matthew Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. 2018. Relational inductive biases, deep learning, and graph networks. *CoRR* abs/1806.01261 (2018). arXiv:1806.01261 http://arxiv.org/abs/1806.01261

[5] Paul C Bauer, Frederic Gerdon, Florian Keusch, Frauke Kreuter, and David Vannette. 2021. Did the GDPR increase trust in data collectors? Evidence from observational and experimental data. *Information, Communication & Society* 0, 0 (2021), 1–21.

[6] Miles Brundage, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, Hyrum S. Anderson, Heather Roff, Gregory C. Allen, Jacob Steinhardt, Carrick Flynn, Seán Ó hÉigeartaigh, Simon Beard, Haydn Belfield, Sebastian Farquhar, Clare Lyle, Rebecca Crootof, Owain Evans, Michael Page, Joanna Bryson, Roman Yampolskiy, and Dario Amodei. 2018. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *CoRR* abs/1802.07228 (2018). arXiv:1802.07228 http://arxiv.org/abs/1802.07228

[7] Moritz Büchi, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, and Shruthi Velidi. 2021. Making sense of algorithmic profiling: user perceptions on Facebook. *Information, Communication & Society* 0, 0 (2021), 1–17.

[8] François Chollet et al. 2015. Keras. https://keras.io.

[9] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences.* Academic press, New York.

[10] Vijay Prakash Dwivedi, Chaitanya K. Joshi, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. 2020. Benchmarking Graph Neural Networks. *CoRR* abs/2003.00982 (2020). arXiv:2003.00982 https://arxiv.org/abs/2003.00982

[11] Upol Ehsan and Mark O. Riedl. 2021. Explainability Pitfalls: Beyond Dark Patterns in Explainable AI. *CoRR* abs/2109.12480 (2021). arXiv:2109.12480 https://arxiv.org/abs/2109.12480

[12] Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli. 2020. A Fair Comparison of Graph Neural Networks for Graph Classification. In *8th International Conference on Learning Representations, ICLR 2020, April 26-30, 2020*. OpenReview.net, Addis Ababa, Ethiopia. https://openreview.net/forum?id=HygDF6NFPB

[13] Facebook. 2021. Facebook Privacy Checkup. https://www.facebook.com/help/443357099140264. Accessed: 2021-11-10.

[14] Yucai Fan, Yuhang Yao, and Carlee Joe-Wong. 2021. GCN-SE: Attention as Explainability for Node Classification in Dynamic Graphs. *CoRR* abs/2110.05598 (2021). arXiv:2110.05598 https://arxiv.org/abs/2110.05598

[15] Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer, and Frank Hutter. 2020. Auto-Sklearn 2.0: The Next Generation. *CoRR* abs/2007.04074 (2020). arXiv:2007.04074 https://arxiv.org/abs/2007.04074

[16] Gerald Friedland and Jaeyoung Choi. 2011. Semantic Computing and Privacy: a Case Study Using Inferred Geo-Location. *Int. J. Semantic Computing* 5, 1 (2011), 79–93. https://doi.org/10.1142/S1793351X11001171

[17] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, 6-11 August 2017 (Proceedings of Machine Learning Research, Vol. 70)*, Doina Precup and Yee Whye Teh (Eds.). PMLR, Sydney, NSW, Australia, 1263–1272. http://proceedings.mlr.press/v70/gilmer17a.html

[18] Yahui Han, Yonggang Huang, Lei Pan, and Yunbo Zheng. 2021. Learning multi-level and multi-scale deep representations for privacy image classification. *Multimedia Tools and Applications* 0, 0 (23 Oct 2021). https://doi.org/10.1007/s11042-021-11667-5

[19] Jun Hu, Shengsheng Qian, Quan Fang, Youze Wang, Quan Zhao, Huaiwen Zhang, and Changsheng Xu. 2021. Efficient Graph Deep Learning in TensorFlow with tf_geometric. In *MM '21: ACM Multimedia Conference, October 20 - 24, 2021*, Heng Tao Shen, Yueting Zhuang, John R. Smith, Yang Yang, Pablo Cesar, Florian Metze, and Balakrishnan Prabhakaran (Eds.). ACM, Virtual Event, China, 3775–3778. https://doi.org/10.1145/3474085.3478322

[20] Bogdan Ionescu, Henning Müller, Renaud Peteri, Asma Ben Abacha, Dina Demner-Fushman, Sadid Hasan, Mourad Sarrouti, Obioma Pelka, Christoph Friedrich, Alba Herrera, Janadhip Jacutprakart, Vassili Kovalev, Serge Kozlovski, Vitali Liauchuk, Yashin Dicente Cid, Jon Chamberlain, Adrian Clark, Antonio Campello, Hassan Moustahfid, and Adrian Popescu. 2021. The 2021 ImageCLEF Benchmark: Multimedia Retrieval in Medical, Nature, Internet and Social Media Applications. *Lecture Notes in Computer Science* 0, 0 (01 2021).

[21] Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* 20, 11 (1998), 1254–1259.

[22] Rui Jiao, Lan Zhang, and Anran Li. 2020. IEye: Personalized Image Privacy Detection. In *6th International Conference on Big Data Computing and Communications, BIGCOM 2020, July 24-25, 2020*. IEEE, Deqing, China, 91–95. https://doi.org/10.1109/BigCom51056.2020.00020

[23] Haifeng Jin, Qingquan Song, and Xia Hu. 2019. Auto-Keras: An Efficient Neural Architecture Search System. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, August 4-8, 2019*, Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis (Eds.). ACM, Anchorage, AK, USA, 1946–1956. https://doi.org/10.1145/3292500.3330648

[24] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, Toulon, France. https://openreview.net/forum?id=SJU4ayYgl

[25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012*, Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Weinberger (Eds.). Curran Associates, Inc., Lake Tahoe, Nevada, United States, 1106–1114. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

[26] Aarthi Suresh Kumar, Anirudh A, Jeet Golecha M, Karthik Raja A, Bhuvana Jayaraman, and Mirnalinee T.T. 2022. Multi Regressor Based User Rating Predictor for ImageCLEF Aware 2022. In *CLEF2022 Working Notes (CEUR Workshop Proceedings)*. CEUR-WS.org, Bologna, Italy.

[27] Junhyun Lee, Inyeop Lee, and Jaewoo Kang. 2019. Self-Attention Graph Pooling. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019 (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, Long Beach, California, USA, 3734–3743. http://proceedings.mlr.press/v97/lee19c.html

[28] Vincent Lefrere, Logan Warberg, Cristobal Cheyre, Veronica Marotta, Alessandro Acquisti, et al. 2020. The impact of the GDPR on content providers. In *The 2020 Workshop on the Economics of Information Security*. HAL CCSD, Online.

[29] Matthieu Manant, Serge Pajak, and Nicolas Soulié. 2019. Can social media lead to labor market discrimination? Evidence from a field experiment. *Journal of Economics & Management Strategy* 28 (2019), 225–246. Issue 2.

[30] Sandra C Matz, Ruth E Appel, and Michal Kosinski. 2020. Privacy in the age of psychological targeting. *Current opinion in psychology* 31 (2020), 116–121.

[31] Van-Khoa Nguyen, Adrian Popescu, and Jerome Deshayes-Chossart. 2020. Unveiling Real-Life Effects of Online Photo Sharing. *CoRR* abs/2012.13180 (2020). arXiv:2012.13180 https://arxiv.org/abs/2012.13180

[32] Aarthi Nunna, Aravind Kannan Rathinasapabathi, Chirag Bheemaiah P K, and Kavitha Srinivasan. 2022. SSN CSE at ImageCLEFaware 2022: Contextual Job Search Feedback Score based on Photographic Profile using a Random Forest Regression Technique. In *CLEF2022 Working Notes (CEUR Workshop Proceedings)*. CEUR-WS.org, Bologna, Italy.

[33] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. 2017. Towards a Visual Privacy Advisor: Understanding and Predicting Privacy Risks in Images. In *IEEE International Conference on Computer Vision, ICCV 2017, October 22-29, 2017*. IEEE Computer Society, Venice, Italy, 3706–3715. https://doi.org/10.1109/ICCV.2017.398

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[35] Tu Minh Phuong, Zhen Lin, and Russ B. Altman. 2005. Choosing SNPs Using Feature Selection. In *Fourth International IEEE Computer Society Computational Systems Bioinformatics Conference, CSB 2005, August 8-11, 2005*. IEEE Computer Society, Stanford, CA, USA, 301–309. https://doi.org/10.1109/CSB.2005.22

[36] Phillip E. Pope, Soheil Kolouri, Mohammad Rostami, Charles E. Martin, and Heiko Hoffmann. 2019. Explainability Methods for Graph Convolutional Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, June 16-20, 2019*. Computer Vision Foundation / IEEE, Long Beach, CA, USA, 10772–10781. https://doi.org/10.1109/CVPR.2019.01103

[37] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 0, 0 (2015), 91–99. http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks

[38] Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, June 18-22, 2018*. Computer Vision Foundation / IEEE Computer Society, Salt Lake City, UT, USA, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[39] Lindsay M. Sanneman and Julie A. Shah. 2020. A Situation Awareness-Based Framework for Design and Evaluation of Explainable AI. In *Explainable, Transparent Autonomous Agents and Multi-Agent Systems - Second International Workshop, EXTRAAMAS 2020, May 9-13, 2020, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 12175)*, Davide Calvaresi, Amro Najjar, Michael Winikoff, and Kary Främling (Eds.). Springer, Auckland, New Zealand, 94–110. https://doi.org/10.1007/978-3-030-51924-7_6

[40] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2009. The Graph Neural Network Model. *IEEE Trans. Neural Networks* 20, 1 (2009), 61–80. https://doi.org/10.1109/TNN.2008.2005605

[41] Eleftherios Spyromitros-Xioufis, Symeon Papadopoulos, Adrian Popescu, and Yiannis Kompatsiaris. 2016. Personalized Privacy-Aware Image Classification. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (New York, New York, USA) *(ICMR '16)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/2911996.2912018

[42] Clemens Stachl, Ryan L Boyd, Kai T Horstmann, Poruz Khambatta, Sandra Matz, and Gabriella M Harari. 2021. Computational Personality Assessment - An Overview and Perspective. https://doi.org/10.31234/osf.io/ck2bj

[43] Liviu-Daniel Stefan, Mihai Gabriel Constantin, and Bogdan Ionescu. 2020. System Fusion with Deep Ensembles. In *Proceedings of the 2020 on International Conference on Multimedia Retrieval, ICMR 2020, June 8-11, 2020*, Cathal Gurrin, Björn thór Jónsson, Noriko Kando, Klaus Schöffmann, Yi-Ping Phoebe Chen, and Noel E. O'Connor (Eds.). ACM, Dublin, Ireland, 256–260. https://doi.org/10.1145/3372278.3390720

[44] Mingxing Tan, Ruoming Pang, and Quoc V. Le. 2020. EfficientDet: Scalable and Efficient Object Detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, June 13-19, 2020*. Computer Vision Foundation / IEEE, Seattle, WA, USA, 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079

[45] Lex Dirk Jan Thijssen. 2020. *Racial and ethnic discrimination in western labor markets: Empirical evidence from field experiments*. Ph. D. Dissertation. Universiteit Utrecht.

[46] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. YFCC100M: The new data in multimedia research. *Commun. ACM* 59, 2 (2016), 64–73.

[47] Ashwini Tonge and Cornelia Caragea. 2020. Image privacy prediction using deep neural networks. *ACM Transactions on the Web (TWEB)* 14, 2 (2020), 1–32.

[48] Christian von der Weth, Ashraf M. Abdul, Shaojing Fan, and Mohan S. Kankanhalli. 2020. Helping Users Tackle Algorithmic Threats on Social Media: A Multimedia Research Agenda. In *MM '20: The 28th ACM International Conference on Multimedia, October 12-16, 2020*, Chang Wen Chen, Rita Cucchiara, Xian-Sheng Hua, Guo-Jun Qi, Elisa Ricci, Zhengyou Zhang, and Roger Zimmermann (Eds.). ACM, Virtual Event / Seattle, WA, USA, 4425–4434. https://doi.org/10.1145/3394171.3414692

[49] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *7th International Conference on Learning Representations, ICLR 2019, May 6-9, 2019*. OpenReview.net, New Orleans, LA, USA. https://openreview.net/forum?id=ryGs6iA5Km

[50] Sergej Zerr, Stefan Siersdorfer, Jonathon S. Hare, and Elena Demidova. 2012. Privacy-aware image classification and search. In *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, August 12-16, 2012*, William R. Hersh, Jamie Callan, Yoelle Maarek, and Mark Sanderson (Eds.). ACM, Portland, OR, USA, 35–44. https://doi.org/10.1145/2348283.2348292

[51] Yibo Zhang, Tawei Wang, and Carol Hsu. 2020. The effects of voluntary GDPR adoption and the readability of privacy statements on customers' information disclosure intention and trust. *Journal of Intellectual Capital* 21, 2 (01 Jan 2020), 145–163. https://doi.org/10.1108/JIC-05-2019-0113