Using AI to classify Instagram's dissident images

PRATA, Didiana; PhD Candidate | FAU USP, Brazil; COZMAN, Fabio; Full Professor | POLI USP, Brazil; POLLETI, Gustavo; Master Student | POLI USP

dissident design, Instagram, data visualization, image classifiers, artificial intelligence

This paper studies the generation of image classifiers, based on artificial intelligence techniques, in the context of research on dissident design in social networks. Given the thousands of images produced and posted throughout the most recent presidential campaign in Brazil, supervised machine learning resources were required to analyse the aesthetic differences amongst posts. An initial sample was captured via tags #desenhospelademocracia; #designativista; #mariellepresente on Instagram. Using visual language parameters, seven aesthetic categories were created in order to classify this massive number of images: factual; memes (subcategory of factual); digital illustration; free-hand illustration; vernacular typography; digital typography and appropriation. Human participation was fundamental in editing, interpreting and improving neural networks during the training process. The results were verified by using a model for training image sets called "MMD-critic", that revealed the subjectivity of the process of interpretation and classifying images using AI. The case study concluded that the data visualization algorithm represents an excellent tool for graphic designers, when artists and curators can handle pixelated images that are analogous to what we see on Instagram. This result goes beyond the mathematical data and algorithms we use when we work with classifying and archiving data; in fact, machine learning opens new perspectives so as to understand the remixed and broad vocabulary of visual narratives on social networks.

Introduction

The emergence of social network user's participation in the production and distribution of images related to socio-political events has led to a new aesthetic, mass communication, dissident design in which citizens, designers, artists and activist collectives participate. Dissident is understood here as the agency strategies that make use of temporary groups, which are organized around common agendas. This paper looks at the creation of aesthetic image classifiers, specially trained to catalog and label, through the lens of graphic design, the message-images introduced by dissident #s (hashtags) on Instagram. The tests conducted on an initial sample of 6,000 images yielded significant results for the field of database image of Brazilian dissident design.

The effort was jointly worked out by an interdisciplinary team of designers and data scientists. The role of the curator and the editor and the creative insights of analysts and programmers are interconnected with the results of the supervised machine learning process.

The social political context in which the images of interest are created, mediated and posted, together with the network's visual culture, the immediacy and the speed of the circulation of facts are crucial for the analysis of the visual language of these image-messages. There is huge aesthetic variety; images are often produced in real time from the appropriation of media images. Some posts are sketchy and poorly made, whereas others are highly elaborate vector illustrations specially created to go around the networks. Given the wealth of graphic experimentation and the poetic potential of these images as a potent graphic memory of Brazilian history – specifically during the election campaign and the first year of the elected president, Jair Bolsonaro's term – new methodologies for collecting and classifying these images were adopted.

There is a new aesthetic vocabulary embedded in the visual language of social networks. Decoding this new graphic language means understanding new ways of producing and sharing images that represent the digital visual culture of social networks. Instagram – the plataform from which images were captured – offers many tools such as photographic filters, various font styles for fun messages, a library of emoticons. The addition of new graphic elements to a photograph is encouraged, facilitated and made available through simple "buttons" for the user to produce their post. The copy-pasting of images circulating on networks and the appropriation of iconic images of illustrious characters add new aesthetic layers to this digital bricolage. The methodology used in this case study were customized to understand these graphic post (message-images) as part of graphic design and communication field.

On Instagram, the possibility of using algorithmic captions (the use of the # plus a keyword) lets one visualize the aesthetics of the database generated by such metadata (Vesna, 2011; Manovich, 2017.) This visualization methodology was used for the selection of visual narratives. We captured images of the following tags: #designativista; #desenhospelademocracia; #mariellepresente; #coleraalegria; #elenao (Figure 1).

Figure 1

Method of development

After choosing which hashtags would be used to set up an Instagram sample, some formal criteria related to the nature, materiality and composition of these images were also defined, as these point to the relevance of these graphic pieces as a new graphic language. The methodology adopted for the classification of more than one hundred thousand images, using Artificial Intelligence and Machine Learning seeks to contribute with new paradigms for the study of the visual language of the images on social networks.

Image training sets using Machine Learning

Image classifiers aim at discerning from a predefined and finite set of labels the one that is appropriate for a given figure. We decided to use machine learning techniques capable of recognizing some aspects of the aesthetic patterns implicit in the visual data themselves and associating them with the labels, thus imitating the eye of the expert. Due to the subjectivity inherent to the multiple visual languages of the dissident images, it was difficult to use an objective algorithm or method capable of performing 100% the classification task. In this sense, we knew from the start that

training procedures for taging images sets depend on the human decisions and have a limited capacity to recognize certains aspects of their remixed visual language.

We also had made a few tests with other computer visual training sets, as Google Cloud Vision and IBM Watson – with commercial licences – using a quite similar learning machine method of training by using visual examples and keywords. Therefore, we were aware of the challenge of creating an aesthetic computer vision labeler.

Cognitive analysis: creating aesthetic categories

In our analysis of this large set of images, we observed some recurring aesthetic patterns inherent to the message-images from the networks. To classify and analyse the aesthetic vocabulary of this material, characterized by the diversity of artistic languages, seven aesthetic categories were established.

These categories were then used for the training of the classifiers, created especially for the classification of the dissident images collected during the election campaign:

1. Factual: Documentary photographs related to fact or character. There may be text interventions. Material characterized by unpolished composition and collage due to the hastiness in which the material is produced to be posted. The main issue is active participation within the network in near real-time.

2. Memes: (subcategory of "factual"): Memes are a subcategory of factual posts. However, there is some elaboration in the message, usually in a more satirical/humorous tone. Memes go viral more easily in the network.

3. Digital illustration: Figurative or abstract vector illustrations made on a specific theme. They represent visual syntaxes. Digital collages with use of photographs also fall into this category.

4. Digital typography: Typographic posts. Digital compositions, made within the application.

5. Free-hand illustration: Cartoons, free-hand drawings, comics, signed or unsigned works.

6. Vernacular typography: Handmade posters, handwritten posts, regionally inspired illustrations, references to vernacular popular culture.

7. Appropriation: posts in which there is an appropriation of the image from a movie poster, a Netflix series, a comic strip, a newspaper page, an artwork or a well-known photograph.

Human participation in the training machine process

The use of machine learning for the classification of this overproduction of dissident images proved to be fundamental for three reasons. Firstly, to assist in the classification of this huge quantitative sampling. Secondly, to demystify artificial intelligence as a "black box" and to examine it as an excellent tool for the visualization of image narratives. And the last and most important reason relates to the subjectivity of the classification activity, be it human or algorithmic. The interpretation of an image by labelers is mathematical. However, the interpretation of the visualization of these images also represents the subjectivity of human classification. The critical points in deciding between one category and another occur both in cognitive categorization by humans and in the reading by neural network classifiers.

Lev Manovich discusses the digital culture and the use of artificial intelligence and provokes us with the following question, "The scale of digital culture demands intelligence that is qualitatively similar to a human, but operates on a quantitatively different scale... How best combine AI with human skills?" (Manovich, 2018, p.34). The author raises the importance of the creative use of supervised machine learning as an excellent tool for designers, artists and researchers. He also proposes a taxonomy to understand the potential of digital culture and the use of AI:

(i) Selecting content from larger collections;

(ii) Targeting content (one-to-one marketing);

(iii) Assistance in the creation/editing of new content, or "participation in content creation";

(iv) Fully autonomous creation (composing music, tracks, writing, creating visualizations from given datasets.)

In the present case study, artificial intelligence is being applied as an assistance in the selection of content based on the seven curatorial classes. We will also point out other AI applications in the creation of future narrative compositions visualized by the classifiers.

Designing "dissident images classifiers" codes

The use of open source scripts of image classifiers has streamlined the process of creating and customizing our "dissident images classifiers". From the data collected and labeled in the previous stage, we initiated the supervised training of machine learning models traditionally used for image classification. Firstly, we implemented the baseline models, logistic regression (LR) and random forest (RF), in Python 3.7 (Bosch et, al., 2007). In practical terms, these models read the pixels of

the images, detecting whether they are composed of photographs, vector drawings, typographies. The recognition is performed by micro pixels, so that a minimum fraction is interpreted, creating horizontal and vertical reading parameters for the configuration of the whole image.

Both LR and RF performed poorly, achieving less than 35% in accuracy. These results suggest that the baseline models underfitted, i.e. They were not capable of effectively capturing the subtle patterns that describe our aesthetic classes. Also, the poor performance of our baseline models points at the need for a more flexible model, such as a deep learning one.

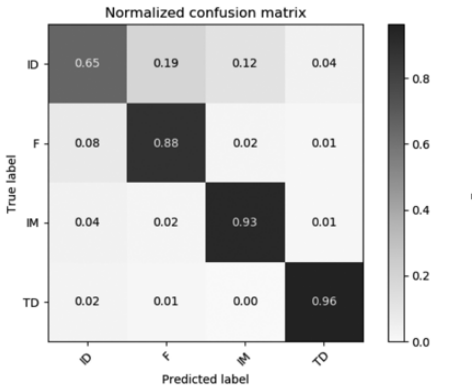
Using Deep Neural Networks

Deep Neural Networks are known for their ability to learn complex patterns, such as in image classification, which makes them appropriate for our case study. These models aim at learning representations from the raw data, so that the algorithm setting and processing stage is performed automatically as part of the machine learning itself. This approach has been successful in a wide variety of domains, notably in computer vision (Krizhevsky et al, 2012; Kaiming et al, 2016.)

The learned representations assumedly capture the relevant information hidden in data so that the classification is carried out using high-level abstractions of the image instead of pixels, for example. Despite being highly accurate, predictions based on these learned representations are modelled in a real valued hyper-space where the dimensions bear no explicit or semantic meaning. Although all deep neural networks aim at learning these latent representations, they differ broadly in how to identify the hidden patterns in data. These models interpret more deeply the reading of images used in machine learning, thus improving the accuracy of the classification.

The convolutional neural net, which is the architecture selected for this work, is designed to consider spatial and locality reference, so that the proximity between pixels is perceived by the model. Thus, CNN models (convolutional neural network) are notably powerful at tackling computer vision, such as image classification (Krizhevsky et. al., 2012), (He, 2015.). We applied it to evaluate the six models of image classifiers by the objective parameters of accuracy. The result can be visualized on the Figure 2 " Visualization of Neural Network confusion matrix". By analysing this table (%), one can note that our model classifies all categories properly, except for Digital Illustration, which is confused with Factual and Free-hand Illustration. After this stage, it became evident that we would be necessary to use interpretability techniques so as to explain why the model is more prone to mistaking one class for another, for example.

Figure 2



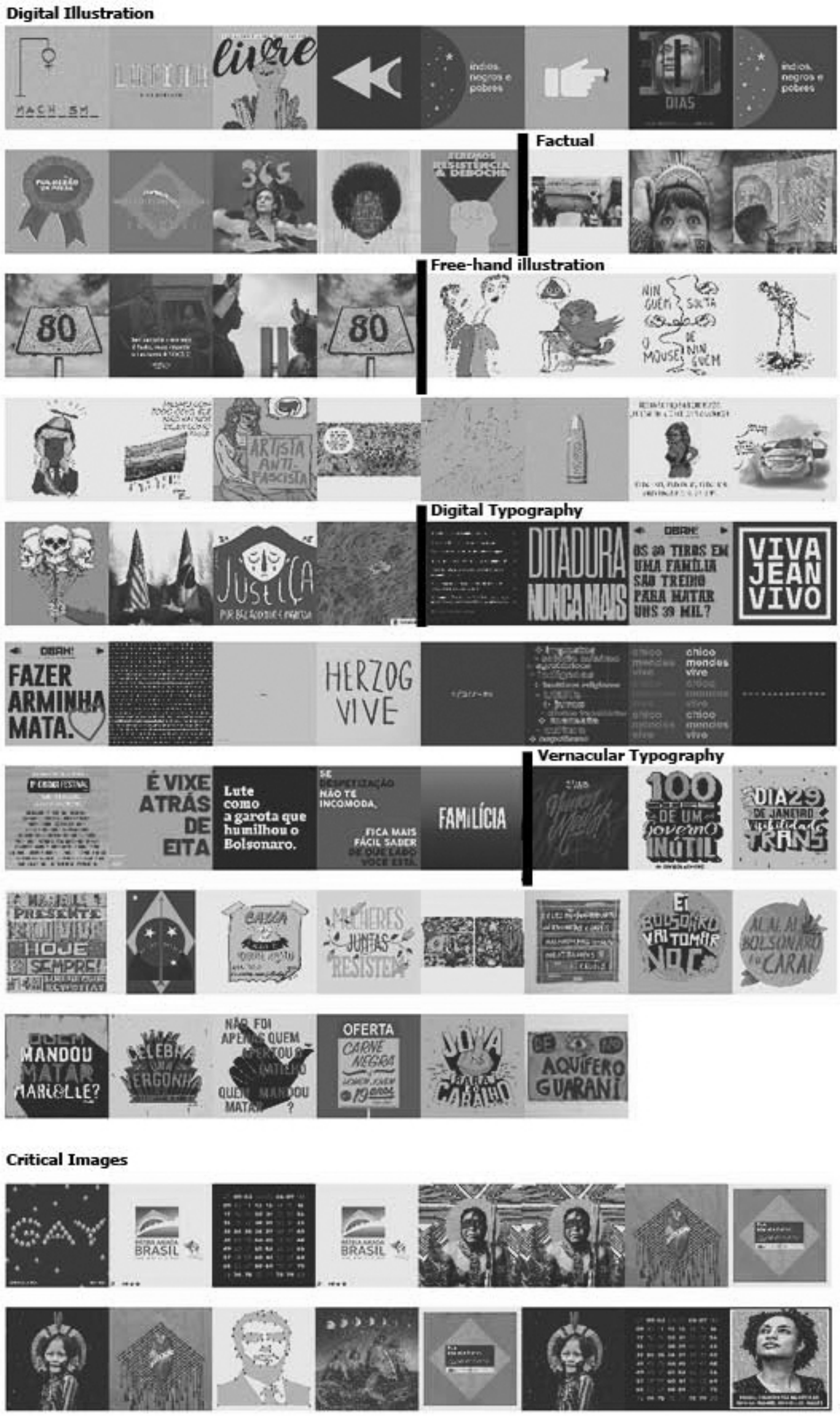
Interpretation of the image classifiers

Classifying large numbers of images to qualify and discover new aesthetic trends is an interdisciplinary activity involving the curator and editor's eye and the programmer's common sense in the training process, in which hits and misses reveal the fragility of the neural network in defining metrics for aesthetic categories. In order to fulfill the need for a more qualitative analysis, we decide to search for a visualization tool that could generate an interpretation of the numerical results, an "AI interpretation image". The visual representation of the quantitative metrics was fundamental in our machine learning training process; for a designer, a sociologist or any citizen who deals with data, tables and numbers are extremely abstract.

We used the MMD-critic model, an open source tool created at the MIT Lab by Been Kim and his team (2016). The "MMD-critic" method aims at identifying simultaneously which images are representative of the whole collection (prototypes) and which ones differ from the rest (criticisms).

In order to analyse the 13% margin of error in the accuracy of the classifiers, we sought solutions for visualizing the classification process. In the sampling used for the training, there were images that could fit into more than one category. In Figure 3 we can see the visualization of the categories with a resolution of 120X120 pixels. The Prototype MMD Critic could recognize and categorize 77% of the images classified by the AI. On the other hand, the Critical images points out that the unidentifiable images belong to the category "Digital Illustration". It is indeed, very complex to recognize images composed of hybrid languages.

Figure 3


Digital Illustration
Factual
Free-hand illustration
Digital Typography
DITADURA NUNCA MAIS
VIVA JEAN VIVO
FAZER ARMINHA MATA.
HERZOG VIVE
Vernacular Typography
É VIXE ATRÁS DE EITA
Lute como a garota que humilhou o Bolsonaro.
FAMÍLICIA
OFERTA
Critical Images

With this setting we were able to detect precisely which images were misinterpreted by the classifier. Interestingly, the "critical" reading images for the neural network are precisely the images in which there is an overlay of techniques: free-hand drawing, cropped photograph, digital collage, handwritten typography. In fact, not even humans would be 100% sure if an image belongs to the "factual" or "digital illustration" category (Figure 4).

Figure 4

The **remixed language** of this image is composed by a cropped photograph of an Indian and worked on together with another photograph of banana tree leaves. In the background, vector graphics allude to ethnic paintings. The original image was interpreted by the editors as a digital illustration because it uses remixing techniques including a documentary photograph. Probably the fact that the photograph is dominant in the image, made it be interpreted as a critical image.

Learning with AI's subjective interpretation

After using this script we decided to eliminate the "vernacular" category (Figure 5). This was a very subtle and subjective category in which we considered regional and cultural aspects to classify its typographic and artistic handwritten style. Hybrid compositions with different materialities result in a remixed language and these compositions could not be detected by the machine. Indeed, by eliminating this category the labelers' performance jumped from 77% to 84%.

Figure 5

For this discussion, the prototypes and criticisms interpretation can provide insights regarding the aesthetic patterns learnt by the machine learning model, thus revealing the "machine narratives." interpreted by the bots.

The classifier training process requires visual acuity and well-defined categorization parameters, large quantitative samples and human assistance at all critical stages. The errors were fundamental for the improvement of the models. The use of these classifiers in the quantitative analysis of dissident images is a major advance in the cataloging of the Brazilian graphic memory.

Final considerations

The case study concluded that the data visualization algorithm represents an excellent tool for graphic designers, when artists and curators can handle pixelated images that are analogous to what we see on Instagram.

Navas says that the visual language of networks represents a cultural remix that is germane to the global digital culture (Navas, 2016). The dilution of authorship in this process of appropriation of different formal elements represents a remixing process. Alluding to the figure of the DJ, used by Navas to describe the "regenerative remix", we can understand how digital images are added from various sources. "Regenerative remix" serves as a bridge to the future of culture in which the ephemeral use of images, text and sound can be used for various creative endeavors. This term is related to connectivity, the use of mobile devices, cellular phone cameras and the constant flow of information exchanged on social networks and the web. In the author's view, "Social media can be considered part of the regenerative remix in terms of discourse. Dissident narratives constantly feed on the flow of images, text and audio from social networks." (Navas, 2016, p.101-105.).

From this concept we can illustrate how these message-images are produced by citizens engaged in activist themes and manifestations. With the same material objects (images, text, videos that

circulate on networks) and the same digital tools (digital illustration software, cameras and scanners on smartphones and the "beautification" and editing functions and filters available in applications), "activist DJs/designers" can produce endless combinations and graphic compositions. And in practice, they add new fragments to the continuous visual narratives, with daily uploads of thousands of posts a day.

Classifying images composed of multiple languages, such as vector illustration collages and photographs and free-hand typography scanned and worked on as digital illustration or a documentary photograph with text messages applied below the image is a complex task and problematizes the remixed language of the networks.

For the machine and for us humans, classification means finding evidence that repeats and sets a recognizable pattern within some parameters. By visually analysing the classifiers' errors it is possible to conclude that the subjectivity of the neural network's interpretation coincides with the editors' classification doubts in establishing the ideal models for training each classifier category.

Due to the ephemeral nature of these images, conveyed in the continuous flow of the networks and doomed to oblivion, archiving and cataloging these images requires a methodology that goes beyond the curator's eye. Facebook and Instagram provide and organize their data algorithms, in such a way that one cannot search for an image by date or by subject. We can track dissident images by user names – not necessarily the author of the image, but by the hashtag preceded by the keyword – in this case the #designativista, #mariellepresente. Or perhaps by geolocation, a resource that was not used in this case.

In this context of precarious archiving of "shared memories" and ignorance of what is actually archived retroactively by companies that store cloud data from social networks, the strategy of creating new parameters for editing and categorizing the production of dissident images is quite relevant. Besides, the creation of specific labelers for this classification has proved to be quite effective.

The use of custom image classifiers for aesthetic, artistic or academic purposes is still at an early stage. Our challenge to design the classifiers in question was to investigate the new visual languages emerging on networks. In this sense, the parameters selected for the training of the classifiers were based on aesthetic rather than ideological or commercial criteria. We could collect and classify a very powerful material, categorized by 6 labels to be explored in other investigations or artistic productions related to dissident images – not only in Brazil, but globally.

The classification process is very subjective since the beginning, from the training set stage. Images classified under the label "digital illustration" are open to new interpretations, for example. They could be "read" as other aesthetic categorizations, as some images have a very remixed language, composed of image overlays and collages. Others are composed of vectorized drawings, masses of colors that make up shapes and outlines. These images inspire different curatorship frameworks, and other graphic reinterpretations.

The qualitative analysis of these images and the recognition of some graphic trends and patterns has been greatly enriched by the quantitative sample, the use of AIs, and the visualization of the labelers' accuracy and doubts.

By visualizing the images sets categories we are able to determine what deserves to be archived in light of graphic design as a new visual language inherent to social networks and as a Brazilian graphic memory. It is up to designers to use new data visualization strategies to our advantage.

In times of "memory of amnesia" and policies of oblivion (Beiguelman, 2019), training supervised image classifiers means editing what is worth being remembered, archived and visualized by the general public, both inside and outside the networks. In this sense, even after demonstrating how it is dependent on and susceptible to human decisions, the use of AI for aesthetic purposes proved to be effective.The interdependence between the image classifiers and the guidelines given by their trainers is a crucial point to understand how the subjectivity of machine interpretation relates to the cultural and ideological repertoire and especially to the main aim of the editor and curator of the images.

In fact, machine learning opens new perspectives so as to understand the remixed and broad vocabulary of visual narratives on social networks.

Figures

Figure 1: Selected images of captured on Instagram by the tags #designativista, #desenhospelademocracia, #mariellepresente, #coleraalegria, #elenao. They represent different aesthetic categories as: factual; memes; digital illustration; free-hand illustration; vernacular typography; digital typography and appropriation.

Figure 2: Visualization of Neural Network confusion matrix (with all categories).

Figure 3: Prototype MMD Critic shows the visualization of the categories with a resolution of 120X120 pixels. The labellers could recognize and categorize 77% of the images classified by the AI. On the other hand, the Critical images points out that the unidentifiable images belong to the category "Digital Illustration".

Figure 4: Post from 4/19/2019, Indian Day, captured on #designativista. The image is made up of multiple languages.

Figure 5: Visually similar Vernacular Images appeared as both "critical" and "prototype" images. Which means that these vernacular and hand written language requires more subjectivity and cultural repertoire to be classified.

References

BEIGUELMAN, G. Memória Da Amnésia: Políticas Do Esquecimento. São Paulo. Edições SESC, 2019.

BOSCH, A. et al. Image Classification using Random Forests and Ferns, in 'ICCV', IEEE, pp. 1-8, 2007.

CRAWFORD, K.; PAGLEN, T. Excavating AI: The Politics of Training Sets for Machine Learning <https://excavating.ai> . Accessed in November 2019.

KIM, B. et al. Examples are not Enough, Learn to Criticize! *29th Conference on Neural Information Processing Systems* (NIPS 2016), Barcelona, 2016.

KAIMING, He et al. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

KRIZHEVSKY, A. et al., ImageNet Classification with Deep Convolutional Neural Networks.In Peter L. Bartlett; Fernando C. N. Pereira; Christopher J. C. Burges; Léon Bottou & Kilian Q. Weinberger, ed., 'NIPS'.pp. 1106-1114, 2012.

LECOUTRE, A.; Négrevergne, B. & Yger, F., Recognizing Art Style Automatically in Painting with Deep Learning.In Min-Ling Zhang & Yung-Kyun Noh, ed., 'ACML', PMLR, pp. 327-342, 2017.

MANOVICH, L. The Language of New Media.Massachusetts. The MIT Press, 2001.

_____. Can we think without categories? 2018 Digital Culture & Society (DCS), Vol. 4, n. 1, 2018, p. 17-28. Disponível em: http://manovich.net/index.php/projects/can-we-think-without-categories. Accessed in November 2019.

_____. AI Aesthetics. Moscow: Strelka Press, 2018 (E-book Kindle).

NAVAS, E. Regenerative Culture. **In: Norient Academic Online Journal.** Available on: <https://norient.com/academic/regenerative-culture-part-15.Accessed in November 2019.

About the author(s)

Didiana Prata is a graphic designer and image curator. Currently she is a PhD candidate at the Post Graduation Program in Design at FAU-USP and Resident Researcher at the Center for Artificial Intelligence (C4AI) at InovaUSP, Universidade de São Paulo, Brazil. Her research focuses on graphic design in different interfaces, database aesthetics and the use of AI as creative strategy. She is also member of the Research Group of Aesthetics' Memory in the 21st Century and Professor of Design at the Faculty of Visual Arts at FAAP, São Paulo, Brazil. ORCID: 0000-0001-6440-3878

Fabio Gagliardi Cozman is Full Professor at Universidade de São Paulo, Brazil, where he works with probabilistic reasoning and machine learning. He has served, among other activities, as Program Chair of the Conference on Uncertainty in Artificial Intelligence, Area Chair of the International Joint Conference on Artificial Intelligence, Associate Editor of the Artificial Intelligence Journal, the Journal of Artificial Intelligence Research, and the International Journal of Approximate Reasoning. ORCID: 0000-0003-4077-4935.

Gustavo Padilha Polleti is pursuing his Master degree on Computer Engineering at Esocla Politécnica da Universidade de Snao Paulo, Brazil. His research interest includes Explainable Artificial Intelligence applied to Recommendation Systems, Latent Feature Models and Conversational agents.