

IMAGE 2TEXT

Starter Pack



Index

Index	2
Introduction: You can do this	3
Section 1: Why are you doing this?	4
Section 2: Ethics as the backbone of every decision you make	6
Section 3: Research your newsroom	8
Section 4: Select a tool	9
Section 5: Set the specific scope	12
Section 6: Criteria for image annotation and tagging	14
Section 7: The tagging games begin	15
Section 8: Keep track of progress	16
Section 9: Implementation	17
Section 10: Share knowledge	18

Introduction: You can do this

Using artificial intelligence (AI) in your newsroom may seem like something out of reach. Maybe you and your team are not computer scientists, developers or have not yet immersed yourselves in the technical aspects of artificial intelligence for image recognition. Or maybe AI-based projects still sound like something that only big newsrooms like The New York Times and the Associated Press have the means to do. We have felt the same way, but now that we have walked this path, developed our own projects (Image2Text, Visión Latina...) and crossed to the other side of the river (although we are of course still learning and will continue to do so), we want you to know that this can be accomplished. Or, in more inspiring terms, you can do this.

Advancements in technology and increased availability of powerful computing resources are enabling the training of large and diverse datasets resulting in more accurate models. Furthermore, new user-friendly low-code tools and platforms are making it easier for non-experts to use and interact with AI systems, which in return contribute to their accessibility.

In the past six months, collaboration has been the most important takeaway from the Journalism AI Fellowship: the more people contribute to AI models, the more diverse and representative the data and results are. We should all create a wide network to share knowledge and help each other. In this sense, we are here to offer our guidance, although we don't have all the answers, we can help you look for them or find the experts that might have them.

Section 1: Why are you doing this?

Set motivations and scope. When setting the scope, it is important to start by clearly defining the problem that the product is intended to solve, as well as the desired outcome or goal. This will help to ensure that it is focused and well-defined.

Once the problem is defined, the next step to begin using image recognition to identify people, objects and contexts in your photos and videos is asking yourselves the right questions:

- Why are you doing this?
- How can this project add value to your newsroom and simplify workflows?
- What ethical considerations should you take into account in order to avoid reproducing biases? For example, if you aim to build a database to recognize politicians in your region, are you making a conscious effort to include women politicians, often undercovered by traditional media?

The answers to these questions will provide you with one of the fundamental pillars of your project: the scope. A modular way of thinking can be useful at this point. You might want to start with a specific module of content: for example, politicians who hold executive appointments in your country during the last decade. After you have achieved that, in another stage of the project, you could expand that same module (to another decade or another level of government, or to politicians that hold legislative appointments, for example) or to another module, such as sportswomen (with an specific criteria, such as “women who played in national teams during the last half a century”).

It might be helpful at this point to learn about the ways others have answered those questions, so we compiled here a few examples:

- [Visión Latina](#) is a project developed by Grupo Octubre, in Argentina, aimed at recognizing political and cultural figures from Argentina and Latin America in journalistic videos and photos
- Panels by Getty Images, a tool for publishers that recommends the best choice of images to accompany a news story.
- Amazon Rekognition offers pre-trained and customizable computer vision (CV) capabilities to extract information and insights from your images and videos.

Once the problem and goal have been established, the next step is to identify the specific steps or tasks that will be required to achieve the desired outcome. This can involve breaking the project down into smaller, more manageable pieces, and prioritizing tasks based on their importance and complexity. It can also be helpful to establish constraints or limitations for the project, such as time and budget constraints, to ensure that the project remains focused and achievable. Finally, it is important to regularly review and adjust the scope of the project as needed to ensure that it remains on track and aligned with the original goals.

Section 2: Ethics as the backbone of every decision you make

As we mentioned before, the best moment to bring ethical considerations into your project is at the very beginning. This starts with taking a look at our own biases, in order to be aware of them and avoid reproducing them into the database we are building. As an example, if my own bias makes me think about men athletes when I write about sports in my region, the database is going to be composed of photos and tags of men athletes and exclude women. If we recognize these types of biases beforehand, we can create a specific criteria to eradicate them.

Furthermore, it is important to ensure that the model is fair and unbiased, and that it does not discriminate against any particular group of individuals. This can involve carefully selecting and preprocessing the data used to train the model, and regularly evaluating the model's performance.

In addition to this, we should always consider the potential impacts of the model on individuals and society, and to ensure that it is being developed and used in a responsible and transparent manner. This should result in conducting regular audits and evaluations of the model, and being transparent about its capabilities and limitations. It is necessary to examine the potential long-term impacts of the model, and to ensure that it is being developed and used in a way that is sustainable and responsible.

It is an obligation to review the composition of the teams designing the models. The diversity of backgrounds, perspectives, and expertise will help to foster

collaboration, representation and innovation, and can also help to ensure that AI is developed and used in a fair, ethical, and responsible manner.

Some useful questions to think about our biases are:

- Is your team diverse? Who are the ones making the final decisions? How can this affect the outcome?
- Are there women, people of color and other underrepresented groups in your team? Are there people from the larger world in your team?
- Is the data used to train the AI model representative and inclusive of all relevant groups?
- Have steps been taken to avoid or mitigate bias in the data and in the AI model itself?
- Are the potential impacts of the AI model on individuals and society being considered and evaluated?
- Is the AI model being developed and used in a transparent and accountable manner?
- Are the long-term impacts of the AI model being considered and addressed?
- Is the AI model being used in a way that respects the privacy and rights of individuals?
- Are the ethical implications of the AI model being considered and addressed?

Section 3: Research your newsroom

Before you start working with your database and training, there is more preparation to be done.

To begin with, you need to think about your users. Who are you building this for? Have conversations with the journalists, editors, photographers that are going to use the tool. Learn about their workflows to create, label and use images. Is there something in those workflows that isn't working for them? Where can they gain efficiency? Which steps of that process are they not willing to change? Having these conversations early on in the process will contribute to creating a tool that they want to adopt and feel their own. And will lead to a better outcome.

You also need to consider what images you want to (and are able to) work with. Go deep into the archive of your newsroom. What is there? What kind of photos and videos, and in what qualities? From what times? Who is portrayed in them? How are currently they being used?

Another aspect that needs to be thoroughly researched is costs. There are open sourced tools and free trials that might be useful for your project, but your newsroom might also need to invest in cloud storage, software and team. The initial tagging process is going to need human effort and supervision. Maybe you already have these resources in your organization, but you might need to hire archivists or journalists. And if your newsroom isn't working with a Media Asset Manager (MAM) still, you might want to consider doing so, to add efficiency to the tool you are creating. You can also benefit from researching potential partners and mentors.

Section 4: Select a tool

Now that you have done your preliminary research, the time has come to choose the tool you are going to use to upload photos and videos, train the model to identify the people that you aim to recognize, and manage your assets. A very big decision in this realm is whether you are going to use an open-source tool or a pre-trained image recognition SaaS. There are pros and cons to both.

Open-Source

Open-source tools' main advantages are that you won't have to pay for a license to use it, and that you will have more freedom to transform it and adapt it to your specific needs.

- These platforms are highly customizable and flexible, allowing users to modify and extend the platform to meet their specific needs (useful for specialized or unique tasks).
- There is a high level of transparency and interpretability: If you need to understand the model's decision-making process this is the way to go.
- Cost: Most open-source platforms are free to use.

However, this will require some technical expertise within your team.

Pre-trained SaaS

Pre-trained image recognition SaaS available in the market such as (AWS Rekognition, Microsoft Azure...) are easier to implement and can be deployed in a short timeframe. However, these have many limitations that should be taken into account:

- Limited customization: Since these are designed to be generic, they tend to be less effective for specific tasks, and they are prone to replicate bias and

discrimination. When testing some available tools in the market, neither we tested were trained to recognise women nor men political leaders from our countries (Paraguay, Argentina and The Philippines).

- Black box models: The inner workings and decision-making process are not transparent or interpretable. This can make it difficult to understand why the model is making certain predictions. This is linked to the point above.
- Dependence on the provider.
- High costs: They usually come with a subscription or per-use cost, which can add up quickly if the model is used extensively. Pro tip: Many have free trials, which you can use to get familiar with their workflows and functionalities, and help you discern if they are the right choice for your project.

Here are some pre-trained image recognition SaaS available in the market:

- Open AI
- Amazon Rekognition
- IBM Watson
- Google Computer Vision

If you are interested in building your own UI, and training your model at your own convenience you might want to look at also some pre-trained models available for computer vision tasks such as:

- ImageNet: A large-scale dataset of annotated images.
- COCO: A dataset of common objects in context, commonly used for object detection and segmentation tasks.

- VGG: A family of deep convolutional neural network models, trained on the ImageNet dataset, commonly used for image classification and other vision tasks.
- YOLO: A model for object detection, which can identify and locate objects in an image.

A couple of best practices for storing images and videos for image recognition

You should also take into account where you are going to store photos, videos and data. In order for the tools to analyze photos and videos and create the associated data, they need to be uploaded to some kind of cloud container. This will probably add costs to your project.

Here are a couple best practices for storing images and videos for image recognition tasks:

- Store the images and videos in a standardized format, such as JPEG or PNG for images and MP4 or AVI for videos.
- Store the images and videos in a centralized and organized manner, using a dedicated storage system or database.
- Regularly backup and secure the images and videos.
- Manage the data carefully to avoid any issues with privacy or consent.

Section 5: Set the specific scope

In order to have a clear and efficient workflow, you need to define who you are aiming to train the algorithm to recognize. As an example, you might want to recognize the main political figures from the past decade. Who falls into that category is open to interpretation. For this reason, you are going to need a specific criteria, such as: presidents, vice presidents, governors, and party leaders from 2010-2020. Our recommendation is setting this criteria with the help of journalists from the specific section you are working with (in the example, politics).

Now you are ready to create a list with everyone that you aim to recognize. You can include their names and some fields for description. Following the example from the previous paragraph, you can have a spreadsheet with: name / charge / start and end of charge dates / political party.

This will be vital for organizing work during the next steps of the process. Once that you have crossed everyone (or almost everyone) in that list, you can broaden the scope and include more names (again, following a specific criteria).

Best practices for data preprocessing for image recognition

Data preprocessing for image recognition is an extremely manual process that typically involves many hours to prepare the images for analysis by a machine learning model:

- Load and format the images: Convert the images from their original format into a format that can be used by the model. Resize the images to a uniform size.
- Normalize the data: Scale the pixel values of the images.

- Split the data into training, validation, and test sets: Use the training to train the model, the validation set to fine-tune the model's hyperparameters, and the test set to evaluate the performance of the trained model.
- Augment the data: Data augmentation is a technique used to artificially increase the size of the training set by applying various transformations to the existing images (e.g. rotating, flipping, cropping, etc.). This can help the model generalize better and improve its performance on new, unseen data.

Section 6: Criteria for image annotation and tagging

Ideally, when you start annotating and tagging (matching detected faces with specific tags), you will want to provide more than one description tag in order to improve the model's performance.

In our specific use case we have found that adding a description of the people we were interested in identifying enriched the metadata available for search and content creation. For example, if your aim is to describe athletes, you might want to add in which team they play, the years that they have been active and if they have won important tournaments. Then comes the names, how you are going to name the athletes. For example, Pelé, the great brazilian soccer player is very well-known by that name, but his full name is actually Edson Arantes do Nascimento. How would you add this information?

Here are some recommendations for image annotation and tagging:

- Use clear and concise tags
- Establish a consistent tagging scheme: Make sure that the tags are applied consistently across the dataset.
- Use a diverse and balanced dataset: It should include a wide range of people and demographics. This can help the model generalize better and improve performance on a wide range of people.
- Test bias correction models to reduce discrimination within your own image tagging process.

Section 7: The tagging games begin

At this point, you have already done a lot of the hard work, and the actual tagging should be easy if you have already created good criteria and defined your scope. The process will consist of uploading the photos and videos that you have gathered and running the analysis with the tool that you have chosen in prior steps. You will get faces recognized as “unknowns”, and your job is to indicate the tool the names of those personalities. Don’t forget to track the process and everyone you have already tagged.

You can run trials by uploading new media of those people and check if they are recognized. In our own experience, the training process doesn’t take a lot of time. Once you have indicated to the system that person X is called XYZ, the next time you upload a picture or video of that person it will already be recognized by that name. But play around with it, try with pictures with and without facemasks, different qualities, different angles and so on.

In many tools you will be able to adjust the accuracy percentage and this will lead to different results. You should play around with these variables, until you find the optimum that works for you and your newsroom.

Section 8: Keep track of progress

As we mentioned on Section 8, it is vital that you methodically keep track of the progress made while tagging. This will help you evaluate if you are reaching your goals on the time that you set yourselves for that and to create new objectives once those are accomplished. You can do this with a simple spreadsheet.

This is an example of the one that was used in Grupo Octubre in Argentina:

PERIODO 2010- 2021				
PAÍS	MANDATARIO	PERÍODO		
Bolivia	Evo Morales Ayma	22/01/2006 – 10/11/2019		
	Jeanine Añez Chávez	12/11/2019 - 8/11/2020		
	Luis Arce Catacora	08/11/2020 - actualidad		
Brasil	Luís Inácio Da Silva	01/01/2003 - 01/01/2011		
	Dilma Rousseff	1/01/2011 - 31/08/2016		
	Michel Temer	12/05/2016 – 31/12/2018		
	Jair Bolsonaro	1/01/2019 - actualidad		
Chile	Michelle Bachelet	11/03/2006 – 11/03/2010 y 11/03/2014 – 11/03/2018		
	Sebastián Piñera	11/3/2010 – 11/03/2014 y 11/03/2014 - 11/03/2018		
Colombia	Alvaro Uribe	7/08/2002 – 7/08/2010		
	Juan Manuel Santos	7/08/2010 – 7/08/2018		
	Iván Duque	7/08/2018 - 7/08/2022		
	Gustavo Petro	7/08/2022 - actualidad		

Section 9: Implementation

The final users of the tool you are creating are the journalists, product managers, photo editors and photographers of your newsrooms. Involving them in the project at every step will help you with implementation, because they are the ones that need to believe in the project and the solutions it brings to their daily jobs in order to incorporate it into their workflow.

For all of these reasons, you should start holding meetings and informal talks with them as soon as you can. You should also listen to their particular needs and be flexible.

Once the tool is ready to be implemented, it might be useful to set specific meetings to show the newsroom how it works and what they can do with it. There can also be some “pilot” time, in which they can play with the tool and get familiar with it before it is officially launched.

Another step that we found useful is to write a short guide with instructions that they can consult. We also recommend keeping open communication channels for users to make as many questions and comments as they need during and after implementation.

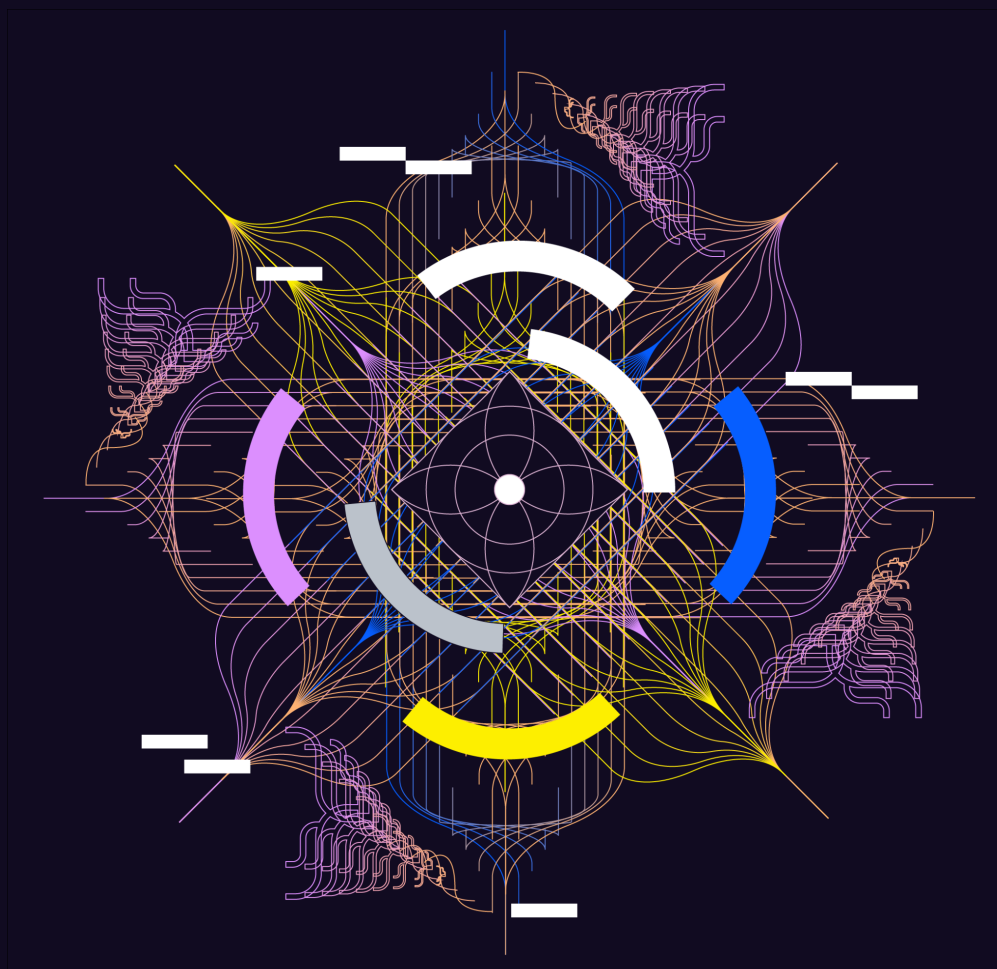
Section 10: Share knowledge

By this point, you have created something very valuable to the journalistic community. You went beyond what is already out there in big databases (that often lack representation) and created knowledge that is both local and specific, and valuable for others. As we mentioned in the introduction, we believe that if more people from different parts of the world and cultures contribute to the AI ecosystem it will become more diverse and representative of different realities. This is an important and urgent goal, since the databases and tools that are being built today will create meaning and have effects in the future.

A big part of the research that we conducted at Image2Text was aimed at figuring out how the knowledge that different newsrooms from different cultures produce can be made collaborative and shareable. And it is a bit tricky. Most tools work in a way that could be defined as “closed”. What you create inside your gates stays there, and you can feed what you are doing with databases created by others, but you cannot feed those databases yourself.

For this reason, we created an open tool, in which anyone can upload media, train the model and the outcomes of that process will be open and collaborative. The next person that uses the model will already be able to use the knowledge that you have created. The downside is that, at least at this stage, its entry barriers are a bit high, in the sense that you will probably need a developer on your team in order to use it. We are working on making it more accessible, without losing the open character that sets it aside from other tools (it can work well as a complement to those!)

Another way of sharing knowledge is speaking to other newsrooms, taking part in workshops and events, having conversations and sharing thoughts, experiences and ideas. We believe that community building is key for a better journalism and better tech ecosystem. Because of this, let's keep in touch :)



This project is part of the [2022 JournalismAI Fellowship Programme](#). The Fellowship brought together 46 journalists and technologists from across the world to collaboratively explore innovative solutions to improve journalism via the use of AI technologies. You can explore all the Fellowship projects at [this link](#).

The project was developed as a collaboration between El Surti, GMA News and Grupo Octubre. The fellows who contributed to the project are: Jaemark Tordecilla, GMA News; Raymund Sarmiento, GMA News; Lucila Pinto, Grupo Octubre; Nicolas Russo, Grupo Octubre; Sara Campos, El Surti.

[JournalismAI](#) is a project of [Polis](#) – the journalism think-tank at the London School of Economics and Political Science – and it's sponsored by the [Google News Initiative](#). If you want to know more about the Fellowship and the other JournalismAI activities, [sign up for the newsletter](#) or get in touch with the team via hello@journalismai.info.