GUIDES AND REPORTS

multimarkts

The 7 deadly sins of AI

The 7 deadly sins of AI

Currently, the conversation about Artificial Intelligence tends to focus on the promotion of technological solutions rather than on its understanding and dissemination. This leads to a mystification of AI, presenting it as an infallible technology capable of automating complex processes and even acquiring consciousness.

In this paper, we will approach AI from a realistic perspective, highlighting its potential but also its limitations. Aimed at developers, Technology Product personnel, CEOs and AI enthusiasts, we explore the 7 "deadly sins", i.e. the 7 common mistakes in its implementation and how to avoid them.

Dani Gómez Domínguez

Director of AI at Multimarkts (Next Chance)

Data scientist with over seven years of experience in Artificial Intelligence and advanced Data Analysis. Working in a startup cultivator, experimenting with the most practical AI possible.

Index

I. Ignorance

Not properly defining the problem to be solved and believing that AI is applicable in all situations.

To talk about Artificial Intelligence, we must first talk about the human intelligence it aims to imitate. In this case, intelligence is defined as the ability to acquire knowledge from environmental information. There are three ways in which humans do this (and not all have been imitated by AI):

1 Deduction

Based on combining known ideas and syllogisms.

Example. f we know that "All men are mortal" and "Socrates is a man," then we can deduce: "Socrates is mortal."

Interestingly, deduction does not allow us to acquire knowledge. It simply relies on combining already known rules of the world, generating new rules. This system was the first to be attempted to be implemented but had little success.

2 Induction

Based on generating rules from experience.

Example: if I only see black crows, I can generate a rule that "All crows are black." If today we see an albino crow, this rule will be invalidated although it will become "Most crows are black."

All that we call AI in the 21st century, the current supervised models need training, a process in which examples are given to the model to generate a new rule through induction. They depend exclusively on experience. This makes it impossible to act generally for all the problems of the world, only for those for which it has been trained.

Example: if I want a model to detect chairs, first I will have to show it thousands of different chairs during training, and it will only work for that. It will not recognize people

③ Abduction

Based on acquiring knowledge through assumptions and probabilities.

Example: If we see the floor wet, we can indicate that someone must have spilled water.

This inference is made casually, without training, and with a high failure rate. In the presence of the puddle, we infer this possibility but we know that it is not the only one: there may be humidity, a broken faucet, condensation from the window... we maintain several explanations at the same time while we find the correct one.

Currently, no AI is capable of imitating this type of logic, which would allow us to escape from the dictatorship of training and have a general artificial intelligence.

> Currently **no AI is able to mimic the logic of abduction**, which would allow us to have a generalist artificial intelligence.
>
> **Daniel Gómez**

If we want to set up our own AI, we will have to think about what type of problem we intend to attack, or how to adapt it to resemble one of these. If we manage to do this, we will have the first step to smoothly set up our own AI. The fundamental problems that AI can solve are:

- **Classification**: Labeling data with one or more categories to which it belongs.
- **Association**: Grouping various data because they resemble each other.
- **Prediction**: Predicting the next data using previous data as a reference.
- **Optimization**: Taking different response paths to certain data to achieve a goal.

Thus, if we want to set up a Netflix-type movie recommendation model, we can approach it in two different ways:

- As an **association problem**, grouping available movies with their tags and description, and returning the most similar movies to the last one the user has seen.

- As a **prediction problem**, trying to predict the next movie a user wants to watch given their previous history.

If our solution is not associated with any of these problems... the solution may not require AI, and there may be other simpler solutions. Or worse, it may be impossible to solve.

II. Arrogance

Thinking that the effectiveness of the AI model will be sufficient to solve any problem.

Every AI model has a failure rate, and it is important to consider the severity of these failures in relation to the problem being addressed.

Example: Severity of a Netflix error recommending movies vs. Error of an autonomous driving vehicle. The latter cannot be implemented even if the failure rate is negligible or much lower than that of Netflix because it would cost us a life.

It is crucial to conduct a **risk analysis** to determine if implementing an AI model is appropriate in terms of cost and benefit.

For this reason, if the failure severity is high, it is important to **change the problem to fall within the safe zone.** This can be done by changing the AI model to reduce errors, either by retraining it or using other state-of-the-art updated models; or by reducing the severity of the error by changing the problem.

Example: we cannot achieve fully automatic driving but we can achieve assisted driving, which helps the human drive and whose failure does not imply an accident.

III. Data Poverty

Not having enough data to train or improve AI models.

The quality and quantity of training data are fundamental to the success of an AI model. It is necessary to ensure that the data is:

- **Sufficient**: depending on the model, we may need more or less data to train it. As reference values, we can have:

Unsupervised models or Zero-Shot

Do not require training data (although their results may not adapt to our solution).

Fine-Tuning

Retraining a pre-trained AI model to adapt it to our solution. As it is already trained, they usually require less data. Usually 1000 data per category.

New training

Training a model from scratch, without any previous training. Usually 100000 data per category.

- **Consistent**: it is important that the data is similar to what the model will encounter during its inference.

  Example: in the model to recognize chairs, it is necessary that the chair images in training have the same quality and size as the chairs it will encounter later.

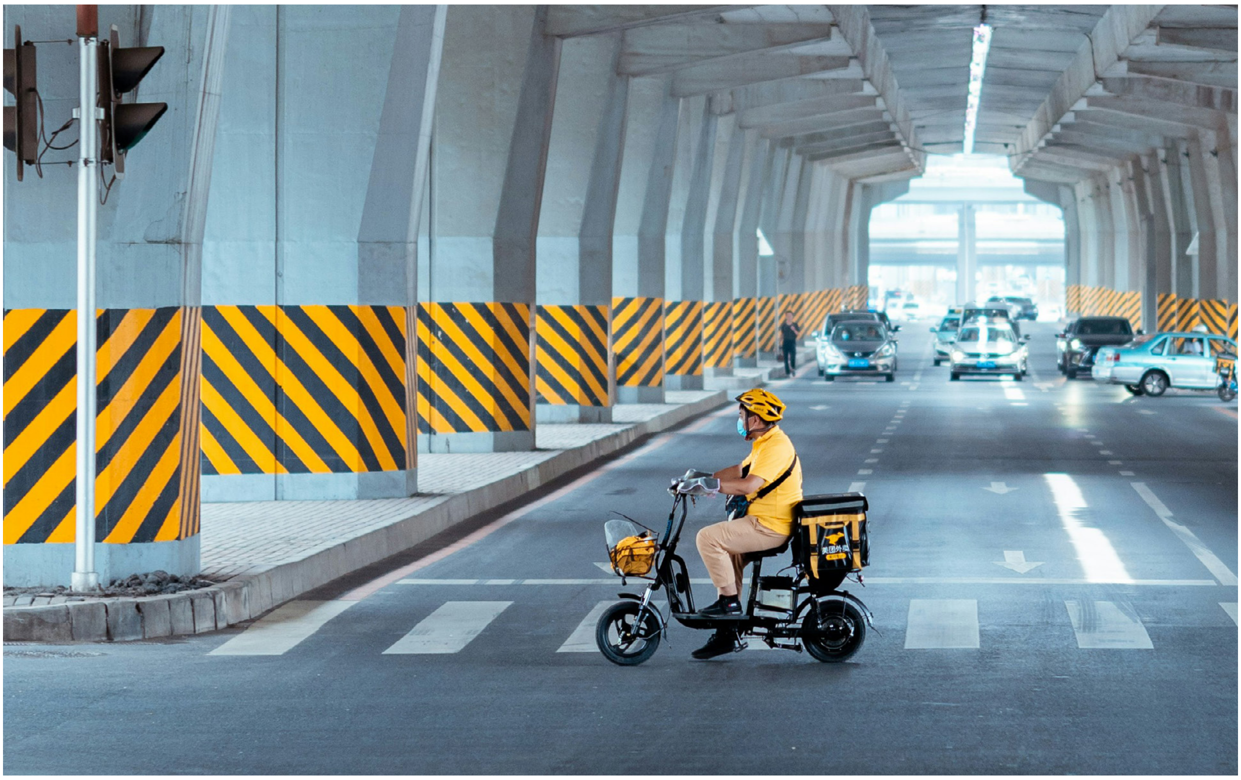
- **Legal**: we can use external databases to train our models, but many have licenses to be used only in research and do not allow commercial or sales solutions.

  It is quite common not to find any public database compatible with our solution. In these cases, we have some last resources that we can follow:

- **Generate our own dataset**: there are open-source tools like LabelStudio to label your training data on your own, there are also businesses associated with this process like Amazon Mechanical Turk. The problem is that this process will require a lot of time and effort.

- **Use synthetic data**: there are studies starting to emerge about the effectiveness of using generative AIs to artificially generate the database. This system is useful and fast but it is dangerous, as the model may inherit biases and flaws from the generative AI used.

- **Dataless alternatives**: as we said, there are Zero-Shot models, based on associations of linguistic and/or visual patterns that allow them to be used without training. The only two problems are that they are usually slow inference models (which makes implementing some solutions difficult) and that they do not always adapt to what we need to set up.

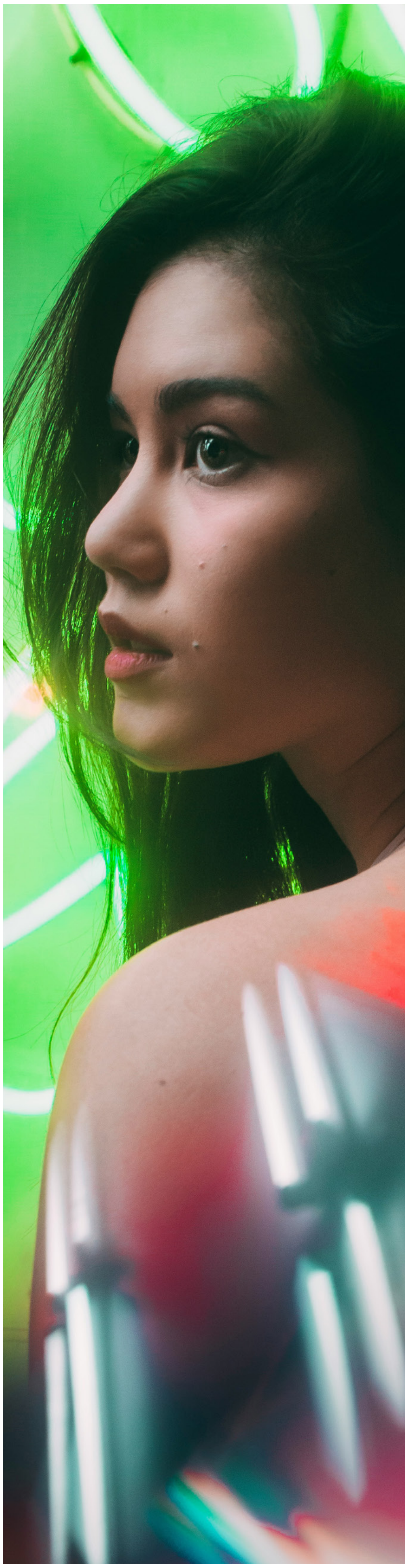


IV. Impatience

Not considering the inference times of AI, which limits some business solutions.

It is important to consider the time it takes for an AI model to return results, especially in real-time applications. We can differentiate AI solutions into two types according to their average inference time:

- **Real-time solutions**: inference takes less than 1 second. Models that meet this condition, except for exceptions, are usually simpler and less accurate, so they offer more possibilities of failure. In addition, solutions that combine several models are more limited because the inference time of all models adds up.



- **Asynchronous solutions:** if the inference takes more than one second, we are facing a solution that must be implemented asynchronously, and the user should not expect the results immediately.

  Example: HeyGen model capable of translating videos, changing the audio to the assigned language, and making it fit with the actor's voice and lip movement on the scene. It took, at least, 3 minutes per second of video entered, a somewhat prohibitive time for a society accustomed to immediate solutions.

Normally large and complex models in demos usually have high inference times, although there are some tricks to avoid it such as:

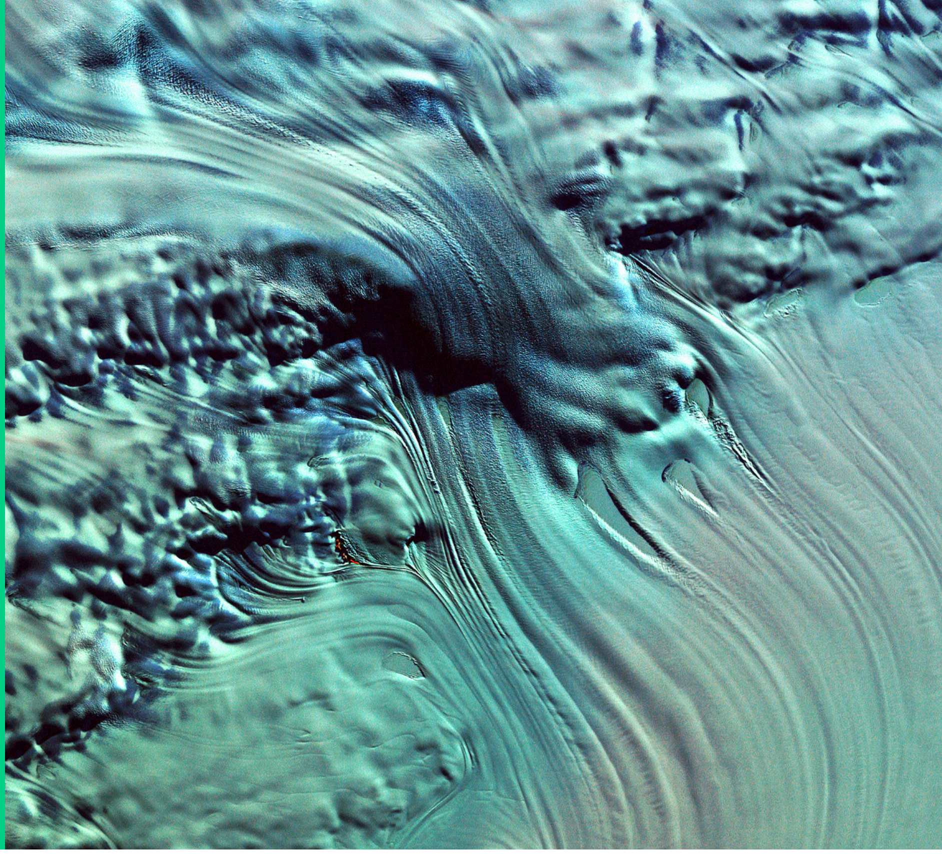
Caching results so they are not called twice.

Autocomplete with similar already generated results to what is requested.

Include something to do while waiting.

Real-time solutions are always more expensive, as they require the model to be available 24 hours a day.

Daniel Gómez



V. Greed

Implementing AI models that are too heavy and costly, difficult to scale.

It is essential to consider the cost of maintaining an AI model on a cloud platform, such as AWS or Azure, including the price of infrastructure and resource consumption. Each machine has its own characteristics regarding power, memory, and price per hour, which we will have to take into account.

Real-time solutions are always more expensive because they require the model to be available 24 hours a day. Also, heavier and more complex models will increase the cost of our solution, as they require more powerful and expensive machines. Another more subtle detail is the number of recurrent calls, if many people call the model at the same time, it will be necessary to hire more simultaneous machines, increasing the cost to double or triple.

In this regard, to implement an AI model we will always have to budget how much it will cost us to maintain that model in the cloud. The cost may be too high for the benefits it can give us, giving a surprise at the end of the month.

VI. Negligence

Not monitoring the quality of the model over time or taking into account variations in its effectiveness.

AI models can deteriorate over time due to the phenomenon of model drifting.

Example: a content recommendation model will not return new movies because they are not included in its training, and its results become obsolete over time.

It is essential to establish evaluation metrics (Example: the number of clicks on recommended content, the percentage of images without detections in a chair detector, etc.) and continuously monitor the model's performance to detect and correct possible problems with retraining.

VII. Obscurantism

Not knowing how the implemented AI model works, using it as a black box.

Although they have complex mathematics underneath, it is not difficult to understand in general terms how a specific AI model works.

It is necessary to understand the strengths and weaknesses of AI models to use them correctly.

ChatGPT and other LMMs are predictive language models, which return combinations of linguistic patterns and keywords associated with given questions. Being a model focused on completing our sentences, it depends too much on the database it has, so it cannot give absolute truths or be used as a source of truth.

Generative AIs create an image, video, or audio using visual/auditory patterns based on what is requested in a prompt. This content is generated using noise as part of its composition, so the results are random and change each time. Faced with rare words, the lack of images in the database causes the results to worsen or be nonexistent.

Image classifiers associate visual patterns with one or several categories through their training, but they do not have to be the most coherent for a human.

Example: in 2016, a model trained to differentiate between wolves and dogs of the Husky species was published. As they are so similar to each other, it was considered that this model could be useful for distinguishing subtle visual aspects. Unfortunately, after several tests, it was found that the model learned to differentiate the snow from the background of the image, not the animal. If there was no snow, it was a wolf. If not, a Husky.

Disclosure and transparency about how models work are essential to prevent misuse of AI.

At Multimarkts, we apply our own AI, taking into account all of the above so that we can offer the best experience to our customers, minimizing errors. We invite you to get to know us and see a real and very profitable demonstration of the use of AI.

multimarkts