ixorthink

UX AND ARTIFICIAL INTELLIGENCE

The importance of user expectations in artificial intelligence projects.

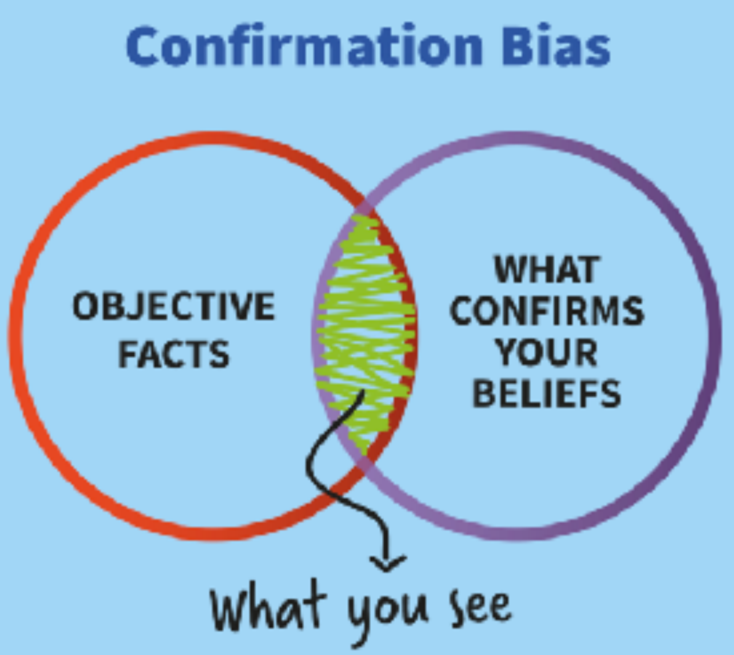
The importance of expectations

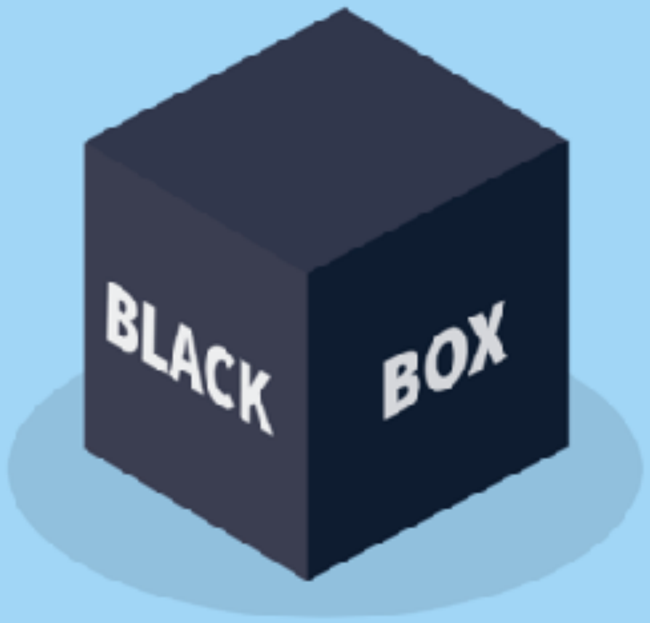
INTRODUCTION

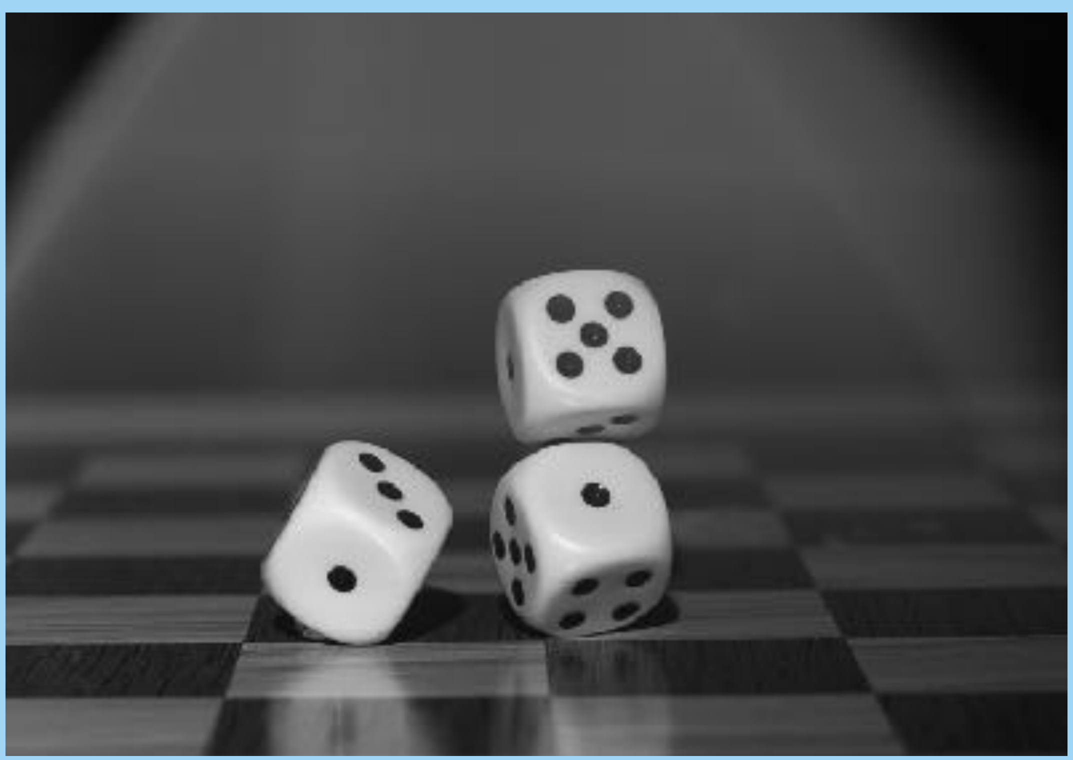
We've all seen movies featuring a super AI taking over the world. Some are convinced that this super AI will become reality, while others may be very skeptical about the possibilities of AI in general. In any case, we could say that all opinions, and therefore all opinions on AI, are never accurate and always biased. These opinions or beliefs shape the experiences we have.

> At election time, people tend to seek positive information about their favourite candidate. When people hear a negative story about their candidate, they will tend to subconsciously ignore this information or interpret it in a positive way. This introduces a bias between people's perception and reality.

This bias is called the **confirmation bias,** which is a cognitive bias that favours information corresponding to one's beliefs. Unfortunately, this bias, that shapes our expectations, keeps us from looking objectively at situations. Moreover it influences the way we experience technology and AI algorithms. Our hypothesis is that the **successfulness of an artificial intelligence application depends on the way we guide these expectations.**







An interesting example to illustrate the confirmation bias is the pilot study at University of California, called **Dice in the Black Box**. For this study a website was created on which a black box AI system was running. The goal of this system was to assess the positive or negative connotation of a users' writing, for example "Your writing is 70% positive". However, the users did not know that the system was a fake AI system. All answers were randomly generated, not using any AI system whatsoever. Surprisingly, over 60% of subjects found this 'random' black box AI system "quite accurate". Participants even tried to explain the occasional mistakes of the AI system by saying "Maybe, I did not give it enough to work with." The study concluded that "**users may place too much trust in a black box system, that is framed as intelligent**".

> "I used 'father' and the score increased, so family related words must be seen as positive."

The confirmation bias also works in the opposite direction: If the participants would be more skeptic about AI, the confirmation bias could have adversely affect their perception. In that case, one single failure is enough to lose trust in a black box system. It is clear that the experiences of a user is almost fully dependent on his or her beliefs and expectations. **If we can pinpoint these expectations, we can also try to adjust them.** This will, in combination with an honest clarification of the possibilities and shortcomings of AI, allow more users to have a **successful experience** with the AI systems we build.

End users' expectations differ from clients' expectations.

In our story of expectations, two different perspectives have to be taken into account. First of all the user can be an end user, who is using our AI application via a tool or web service. Secondly, the user can be a client or stakeholder with a certain business need, who can be met with an AI application. Both types of users operate with different expectations.

We will begin with focussing on expectations of the clients and stakeholders. In this case we want to make sure that we know their expectations from the very beginning of the project. In the second part we will look at an AI system from the perspective of an end user and his user experience.

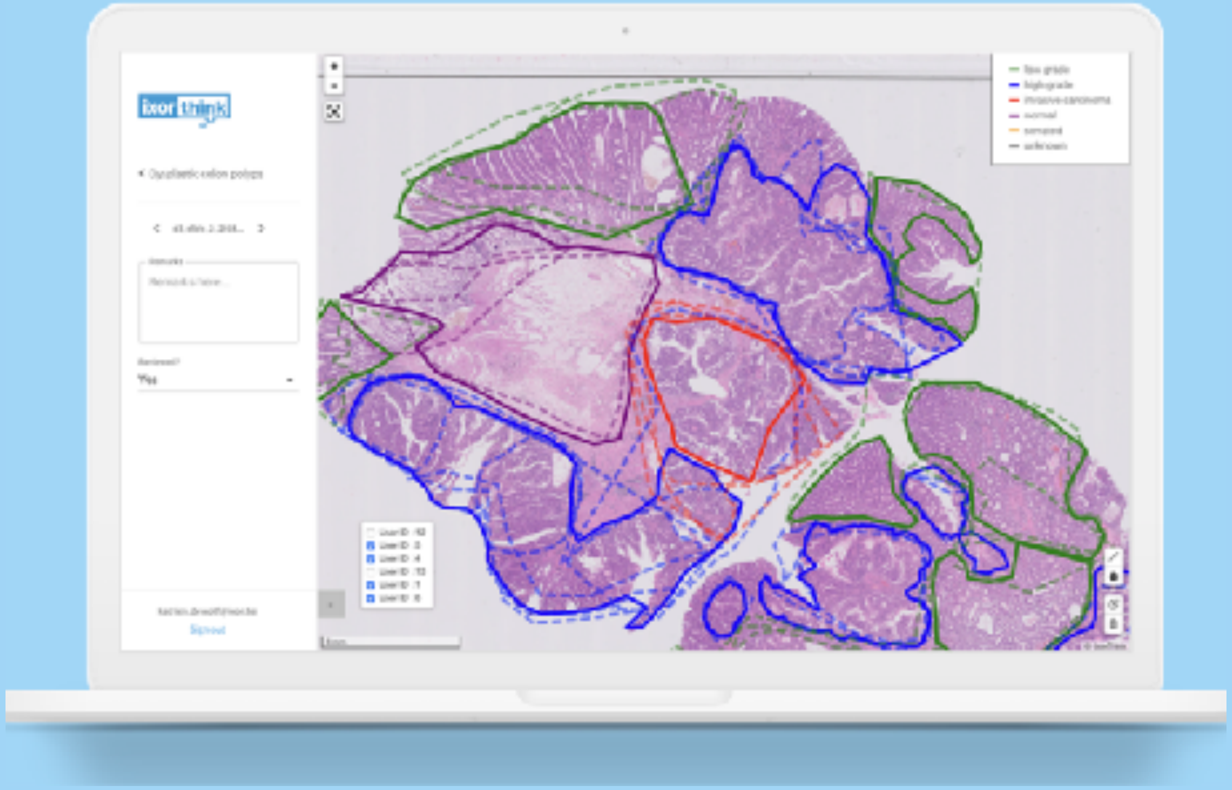
Aligned expectations

CLIENTS & STAKEHOLDERS

Clients have data available but they often don't know in which form it is usable for machine learning. For them, it is not common knowledge that **annotated data or metadata** is in most cases a necessity if we want to learn and generalise from data. When annotated data is lacking, a method should be devised to convert the knowledge of domain experts into the digital form.

In one of our projects in the **health care industry**, the IxorThink platform was the main interface to create data annotations. The aim of the project is to classify the grade of dysplasia in colon polyps. Dysplasia in colon tissue is symptomatic to **colon cancer** which is the third leading cause of cancer-related deaths in the United States. In Belgium, people over 50 are invited to a free screening for colon cancer. Depending on the results of this screening a colonoscopy is performed. During the colonoscopy suspicious polyps are removed. These polyps are then analysed by a pathologist. For this analysis, the tissue sections are made and scanned at a microscopic level. These slide scanners create multiple zooming layer scans of over 4 gigabytes. These scans were annotated by several pathologists, using our **IxorThink platform**. To do the actual classification using AI, the scans are divided into small tiles which are all individually classified by a neural network.

The first results of our machine learning model were disappointing, which is why we decided to review the created data. **These data review meetings were of utter importance to shape our client's expectations.**



BIASES IN DATASETS AND ITS INFLUENCE ON RESULTS

Reviewing the data together with domain-experts made clear that the poor results could be attributed to two unexpected problems.

First of all, some specialists made a more fine-grained annotation than others. It became clear that they needed to decide on specific guidelines to create annotations.
Secondly, there were discussions on the correctness of the annotations. The main reason being that their opinion was shaped by their individual expertise and how they attribute more weight to certain abnormalities. These small differences may lead to a different classification for some parts of the tissue. Clearly, our experiences shape the decisions we make. This kind of bias is the result of the **availability heuristic:** people overestimate the likelihood of events based on their own recent memories. Our solution is to minimise the effect of this bias by making sure that annotations are created by multiple experts together.

Organising **data review meetings** in different phases of an AI project, helps to correct expectations by showing the client how biases, asymmetries, etc. in a training dataset influence the resulting machine learning performance. Biases can never be removed entirely, but **understanding a dataset and its biases is certainly one of the keys to successful AI implementations.**

The development of an AI application is an iterative process, driven by the feedback of business stakeholders.

While the data is clearly the most important aspect, it is not only data understanding that shapes the client's expectations. In practice, most people have a very abstract understanding of machine learning models. Hence, it is of crucial importance to communicate clearly with business stakeholders and make them understand why one use-case is more difficult than another. In general, there is a certain methodology to make sure that you don't promise your clients the moon. By splitting the goal in small phases based on technical difficulty and data availability, we can test results iteratively. Secondly, each iteration should start by sitting together with the stakeholders, all of them with very diverse backgrounds. From this first moment, people with technical knowledge can create a realistic image of the possibilities.

Transparency and trust

END USER PERSPECTIVE

This brings us to the next step. Let's imagine looking at an AI system from the end-user's perspective. When the results of an AI application are presented to the user, his expectations are already influenced by the way we present the results. The user experience is shaped by two big questions:

WHAT ARE THE RESULTS?

The problem here is that that results can be formulated and interpreted in different ways. The output *"This part of tissue is normal with 80% confidence"* is not formulated in an intuitive way, as different users might interpret this confidence differently. A better formulation method should be available to reduce the users' doubt.

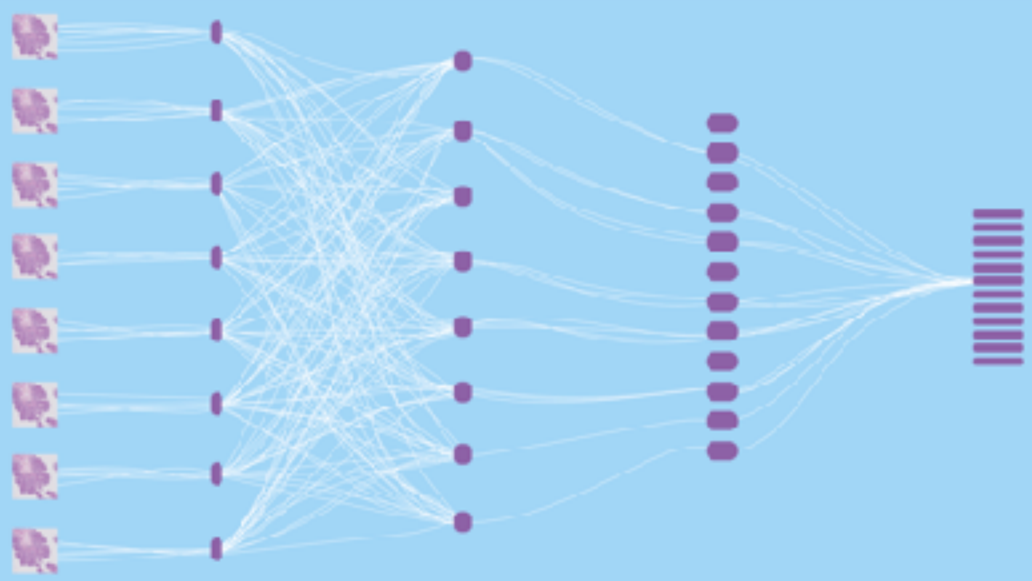
WHY DO WE GET THESE RESULTS?

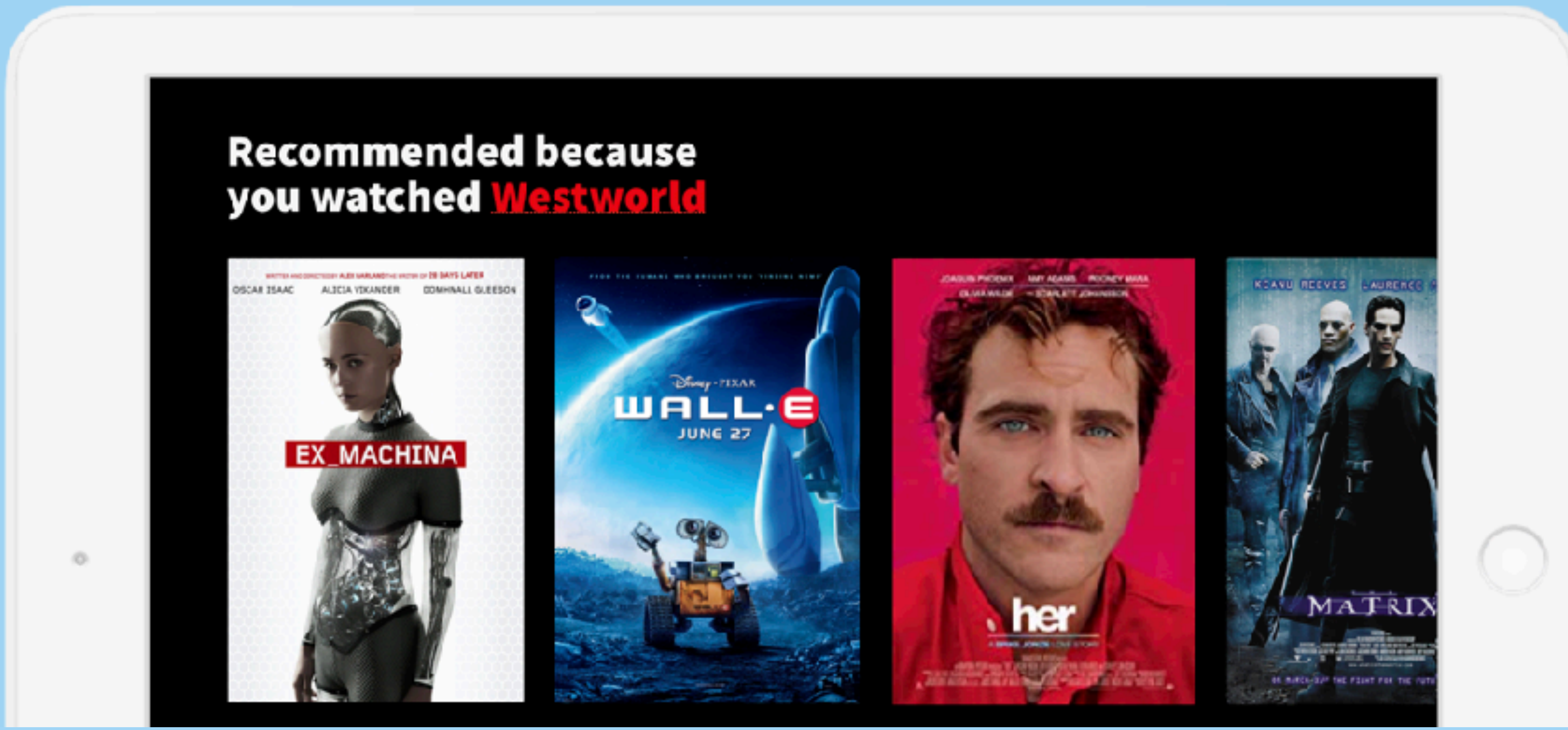
Machine learning is definitely not known for its interpretability. Some users might have a hard time using a system, that doesn't offer a look under the hood. But how can we make sure the AI is not seen as a black box? Well, the opposite of a black box would be Explainable AI or transparent AI.

EXPLAINABLE AI

Explainable or transparent AI is an intelligent system which can be easily trusted and understood by humans. One problem with explainable AI is its definition. What is meant exactly by explainability, and how transparent should a model be?

From a technical perspective, the feasibility of explainable AI is highly dependent on a **model's complexity**. A decision tree model is by definition easily explainable, while complex models like neural networks do not have this property. Some tools and frameworks exist, like LIME, which help to trace results back to the input data. These methods certainly create the possibility to check if a model is properly trained in specific situations. However in many use-cases, like our colon classification project, tools like this do not provide useful feedback. In these cases knowing which exact pixels of the input were important for a classification, is not a reason for additional trust. This is important to notice because the reason to create transparency is first of all to **increase a user's trust.**



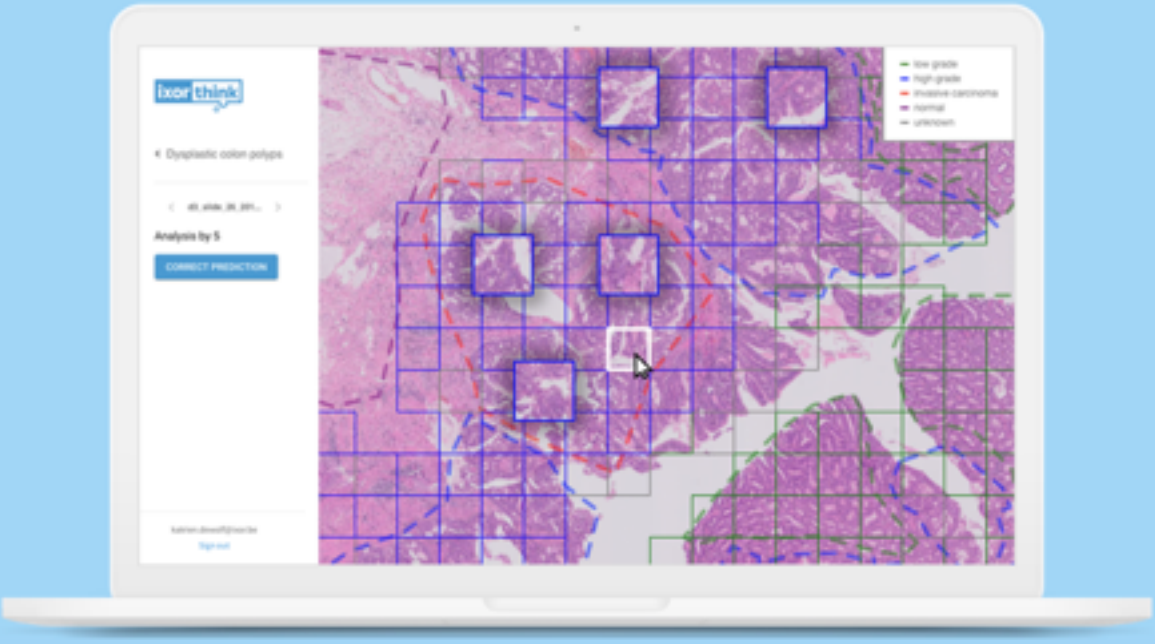


A well-known example of non-technical explainability: in the Netflix interface a lot of recommendation are shown to a user, but these are always explained.

Not all models should and can be made fully transparent. This would conflict with the very reason we use a computer to learn things: to solve complex tasks which can not be solved by using simple instructions. In our day to day life, trust is created by experience. Therefore the only way to increase trust in an intelligent system is by **extensive testing** in the real world.

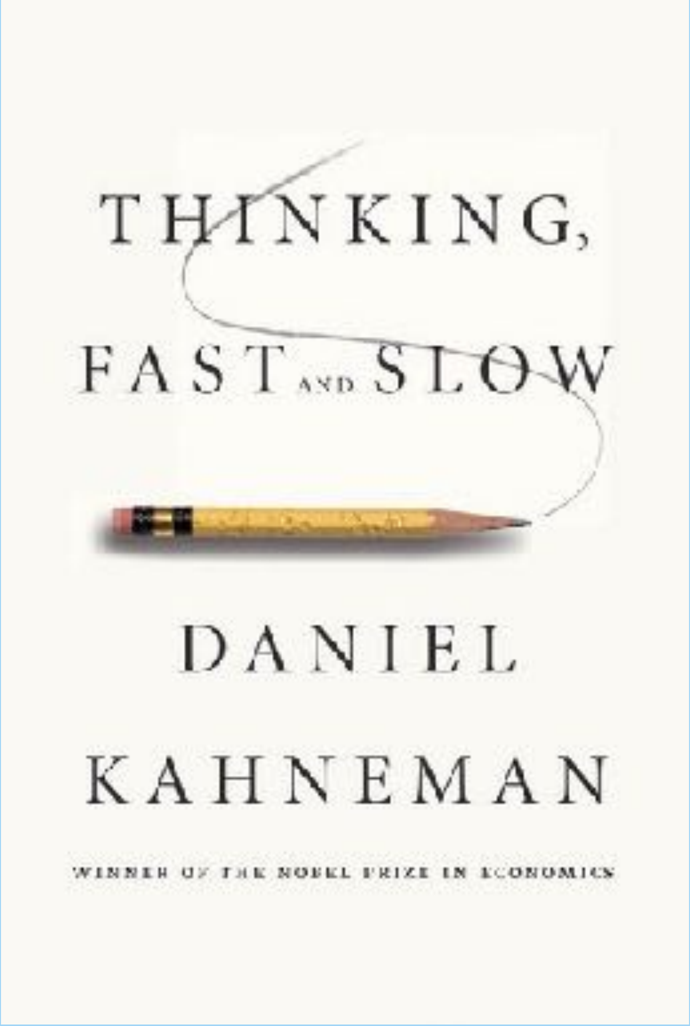
However, the hunger for explainability is still there. In order to feed it, we can try to incorporate explainability in a less technical way. For an end-user it is enough to vaguely understand how results were generated. We call this **Honest AI**. Providing this kind of explainability does not have to be technically difficult. In most cases it is enough to find a way to trace results back to the dataset, instead of trying to explain the exact decision logic of a complex learning model.

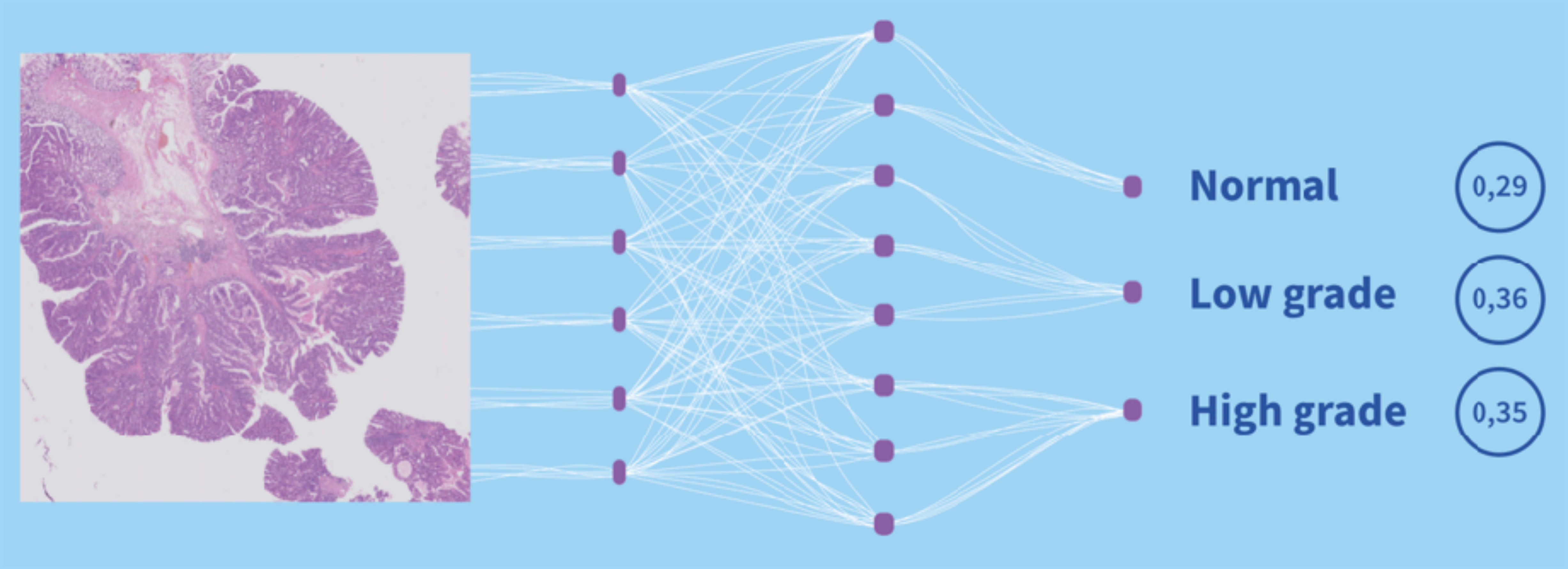
To incorporate transparency into our tissue classification system, we use tile-features from inside the neural network to highlight similar tiles. Showing tiles with similar features helps the pathologist understand why classification are made by the AI. This way we are able to remove part of the feeling of a black box, without the explanation of its inner workings which are too complex to understand.



USER INTERPRETATIONS

Apart from explaining to an end-user why the algorithm spits out this result, there is still the problem of interpretability. In machine learning projects, results are almost always tracked by probabilities, confidences and statistical metrics. However, when we take such project to the outside world, the way we output results requires a second thought. In the book "Thinking fast, thinking slow" Nobel-prize winner Daniel Kahneman explains the **Prospect Theory** which describes that people can not look objectively at situations that involve risk and uncertainty. Even more, it is proven that preferences of individuals are different depending on the way the options are presented.





Output confidences of models should not be seen as real probabilities, because they are merely a confidence estimation based on the data.

To solve our problem, a bolt solution would be to **avoid outputting probabilities**. This way, the decision is made by the application itself, instead of delegating this to an end-user.

When an image scan is analysed by our neural network, the network computes classification confidences for every small tile of the scan. Tiles of the scan tissue where confidences are close together, stay undecided and are tagged as "unknown". By having elaborate conversations with the domain experts, this assessment, choosing when a tile is tagged as unknown, can be done more objectively, compared to when an end-user would have to make the decision himself. This will overall result in a more **objective decision process**.

KEY TAKEAWAYS

A close **collaboration between ML engineers and domain experts**, is essential to improve the results of your AI model, as it helps with **exposing hidden biases in the data**. Therefore it is important to develop and test the model in an iterative fashion.

Having a good model doesn't guarantee a successful project. An important factor that comes into play is **the guidance of the end-users' expectations.** This part of the project is less obvious and should be a phase on its own during the UX design. Having a close **collaboration between UX designers and ML engineers**, will lead to a more objective user experience.