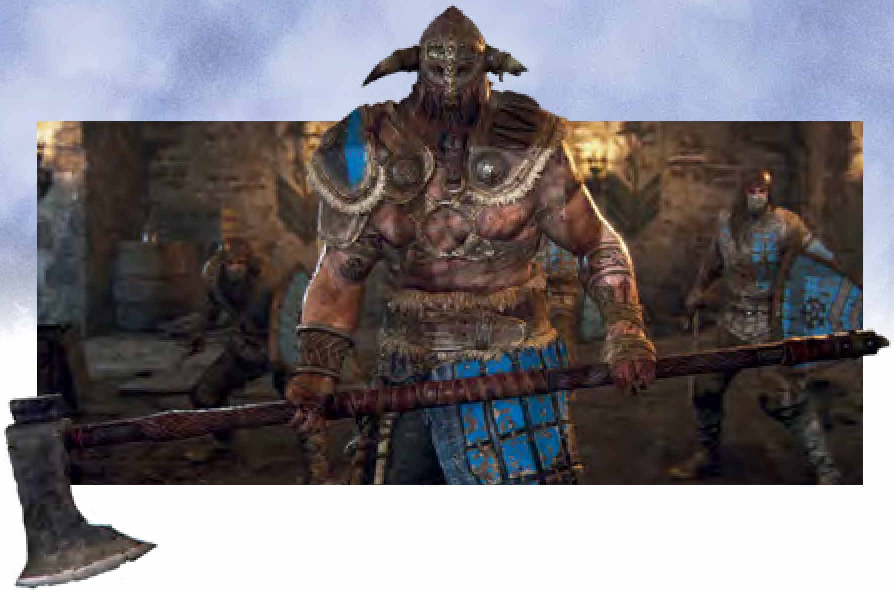
FIGHTING TOXIC ONLINE GAME BEHAVIOUR

For Honor gameplay screenshots
Images courtesy of Jason/Flickr.com (top) and Nick72 Italy/Flickr.com (opposite)



Author: **David Mizzi**

If you've played video games online before, you're probably already familiar with the toxic behaviour found in some online communities. But what if there was a more effective way to moderate online games? ***David Mizzi*** *speaks to* ***Prof. Georgios N. Yannakakis*** *about his latest research with* For Honor.

You sign in, ready to join the multiplayer server and immerse yourself in a friendly online game. As the opposing team pushes back, tensions are high. Suddenly you start receiving flak from one of your teammates — aggressively calling you out on your mistakes and harassing you with insults. You try to mute the player, but your game and fun lie ruined.

For many gamers, joining an online multiplayer game means enduring toxic behaviour. Others avoid the online space entirely and stick solely to single-player games. For numerous video game publishers, encouraging players to enjoy online experiences means directly tackling the problem of toxic online behaviour. However, the problem isn't limited solely to game publishers; it affects a swathe of institutions ranging from voice-chat applications such as Discord to law enforcement. Teaming up with Ubisoft, the world-renowned French video game company, the Institute of Digital Games at the University of Malta have developed an AI to help identify and combat toxic behaviour. As a key player in the industry, Ubisoft's focus is on implementing a rapid and solid reporting system to help encourage prosocial behaviours.

REPORTING OFFENSIVE BEHAVIOUR

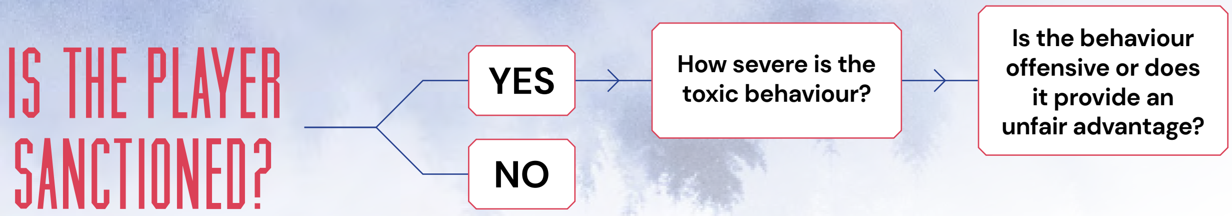
In the vast majority of online games, toxic behaviour is moderated through community peer-reporting. While a good initiative, this approach presents a number of limitations, the biggest one being that it is up to the community to report toxic behaviour. Oftentimes players do not even bother reporting transgressive behaviour, leaving the perpetrators unpunished. This could be because of the effort involved in reporting, reports not being effective, or simply that toxicity has become a normalised part of the player experience.

Some online communities in particular, such as MOBAs (Multiplayer Online Battle Arenas), are notoriously toxic. This could be because of their inherent competitiveness or anonymity. One such game from Ubisoft, *For Honor*, an online battle arena which has players decking it out as Knights, Vikings, or Samurais, formed the basis of the Institute's research.

In 2020, Ubisoft reached out to Prof. Georgios N. Yannakakis from the Institute of Digital Games to develop a solution that could complement community reporting. 'Partners from Ubisoft and other universities reached out to me and said we have all of this rich data, and we would need some help to process it,' Yannakakis explains. 'This isn't the first time the Institute has worked with the industry,' he grins. Yannakakis, working alongside a team of international experts, wanted to see if it was possible to identify toxic behaviour simply by observing in-game behaviour.

DO TOXIC PLAYERS PLAY DIFFERENTLY?

The data provided by Ubisoft presented the team with a dataset of almost 1,800 sanctioned players. Sanctioned players refers to players that have engaged in toxic behaviour and been reported. This was compared to unsanctioned



IS THE PLAYER SANCTIONED?

YES

NO

How severe is the toxic behaviour?

Is the behaviour offensive or does it provide an unfair advantage?

players to create a sanction matrix which organised players according to the severity and type of toxic behaviour.

Through this, the team realised it was not only possible to distinguish sanctioned from unsanctioned players through their in-game behaviour, but it was even possible to predict the sanction severity and type. Essentially it is possible to identify a toxic player based on the types of matches they play (a custom, ranked, or tournament match), how many matches the players abandoned, movement (whether they stand still, walk, run, or sprint), match performance, and chat actions.

'These characteristics were the result of careful analysis. They list the characteristics on one side: whether players are aggressive to their teammates, the frequency of their chats, and how they play. Not all of these characteristics were relevant, but we wanted to see what behaviours could correlate to toxic behaviours,' clarifies Yannakakis.

To put this into perspective, sanctioned players are more likely to run, less likely to play practice matches, have a lower score (wins), and tend to play significantly more in vs AI modes (a type of online match where the player fights against bots).

While organising the data is (relatively) straightforward enough, the next challenge is training an AI to distinguish and predict toxic behaviour. This is where the data science and machine learning aspect comes in, Yannakakis' area of expertise. To do so, the team used a Random Forest Model (RFM).

SEEING THE FOREST FOR THE TREES

'RFM is a good, old-fashioned machine learning method,' explains Yannakakis. In its simplest form, RFM is a series of if-then rules with multiple paths. These multiple paths create multiple decision trees to help process the data. Rather than trying to predict something at once, the decision is split across different 'clusters'.

The first step was to see whether the random forests (the set of decision trees) could distinguish between sanctioned and unsanctioned players within a data set. This is a binary (yes or no: whether a player is sanctioned or not) classification task the random forests needed to decide about, based on the data fed. On average, it was able to do so with a 95% success rate on unseen players!

The next step was to predict the severity of the toxic behaviour: whether it merits a warning or a ban. The model was able to accurately predict this 85% of the time, with a 95% confidence interval across 100 runs lying between 84 and 85%. This means that in 95 out of 100 tries, the program would be able to predict toxicity correctly with an accuracy between 84 and 85%.

The final — and most crucial — step was to see whether the random forests could predict the type of toxic behaviour. Splitting toxic actions into offensive behaviour and unfair advantage, the model was able to predict this with an 87.5% accuracy on average, with a 95% confidence interval lying between 86.7% and 88.5%.

It is important to note that the reported results are on unseen data. The RFM was trained on 80% of the data provided by Ubisoft, and then it was tested on the remaining, unseen 20%.

The beauty of RFM is that it's an 'expressive AI' method. While deep neural networks (the most popular AI algorithm nowadays) present a slog of data as millions of billions of parameters, RFMs present the data in a more accessible, human-readable, manner. 'This makes it easier for the entire production team to consult with it. So the level designers can understand and use this information when designing levels, programmers when coding, or the writers when penning dialogue. It is a transparent model,' explains Yannakakis.

A NEW DAWN FOR ONLINE GAMING

The outstanding success of this research promises to create a more positive online gaming experience. However, it does raise a fascinating ethical question. If we can predict player behaviour with a high degree of accuracy, should we be preventing transgressive behaviour before it even occurs? And if it is possible to predict player behaviour in a game, can we also by extension predict people's behaviour in real life?

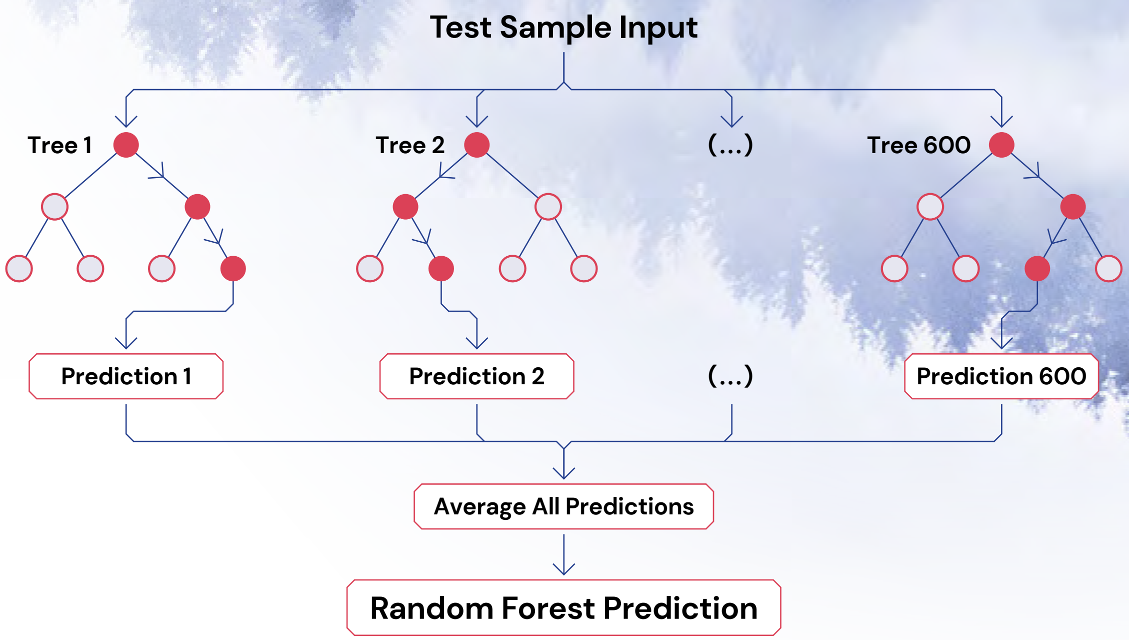
The study makes it crystal clear that the algorithm should supplement the manual efforts of community managers, rather than become a completely automated system. Final human verification is ultimately necessary to fairly impose sanctions.

'We are after a complementary approach. If you simply automate the entire system, then who controls the predictions? You end up having this dystopian system where the AI decides who is toxic and who isn't. The last thing we want is for the AI to have complete control,' laughs Yannakakis.

The next generation of online games may very well utilise the findings of Yannakakis and the rest of the team. With a little luck, online warriors might soon fight with honour in a safer, non-toxic environment!

Prof. Yannakakis is also the current Editor in Chief of the IEEE Transactions on Games, the top-tier journal for technical games research: https://transactions.games

RANDOM FOREST MODEL



For Honor gameplay screenshots
Images courtesy of Nick72 Italy/Flickr.com (middle, bottom right) and Badass Dream/Flickr.com (bottom left)