

Module 2 Exercise¹

Directions: *read the following hypothetical case study; then answer the questions that follow.*

Case Study

Background

In 2020, the gaming market generated almost 200 billion dollars – but even more striking is the large amount of data generated by today’s games. Around the world, there are over 2.5 billion gamers, generating over 50 terabytes of data each day. Today, major gaming companies have all the data they need to understand who their users are and how they engage with the product. As Artificial Intelligence (AI) has entered into the gaming industry, this has only become more pronounced. AI is currently used at nearly every point in the gaming process: from targeting new potential customers to integration into the development of the game itself. These AI tools have allowed for a more immersive gameplay experience, while giving game developers adopting these tools better insight into user behaviors, tendencies, preferences, and other factors of interest. These analytics have become more powerful and have branched out into other forms of AI-enhanced monetization (such as targeted advertising). Some real-world cases have shown that many critical metrics are monitored and used to establish baselines of standard behavior patterns, which can then be used to alert teams or other interested parties to anomalies that might create issues or opportunities. The potential for misuse is concerning. The following fictional (but not unrealistic!) case study further explores this topic.

Case

Zark was a young and ambitious game developer who, after just 8 years, had risen his way to the top at a major gaming company *End* – where, like many other successful gaming companies, they had been monitoring and recording user behavior to see how their games were being played, using those insights to tweak gameplay and improve game experience. When Zark first joined *End* in 2008, the company was strictly using the information to make a variety of in-game changes – such as enhancing storylines, adjusting difficulty levels, and justifying new content updates. Over the years, they developed a remarkably powerful set of tools providing insight into what motivated users, what stopped a player from achieving a goal, what turns them off, and more. These behavioral insights allowed them to become more strategic and savvy, helping them to become one of the most successful gaming companies.

¹ Prepared by Daniel Rosiak, Inamori International Center for Ethics and Excellence.

At first, *End* had been using AI simply to improve the overall gaming experience by making their games smarter and more realistic. But as they began to fully integrate AI into all aspects of the framework of their games, it became clear that mining such metrics not only let them create content better tailored to their users, but often gave them uncanny insight and predictive powers regarding their user's behavior and choices -- in some cases, as they would later discover, even some behavior and choices made beyond the game!

Around the time Zark assumed the position of CEO at *End*, in 2016, he specifically realized that, in addition to the heavy amounts of data collected from users during gameplay, the company had (largely inadvertently at first) been collecting more and more personal data: they were collecting data not just from product telemetry (players' in-game behavior) and psychometric evaluations conducted in-game, but also from external sources. In particular, they had been collecting data and personal info from its users' cellphone-linked profiles, and the AI analytics had been running on that data as well. Paired with this personal data, choices made in-game that revealed things about the player's personality – like dialogue choices, how quickly they give up when presented with challenges – got recorded and stored, helping the developers at *End* isolate commonalities and personality types.

As Zark settled into his CEO position, he made a point of emphasizing that while this information could continue to be used to improve their games, it could also be used to build extremely robust personal profiles of its users, especially when combined with other types of information they were already collecting. While Zark and others at *End* immediately saw the profit potential for such profiles – especially to advertisers and various internet applications – and had intentions to pivot the company away from game development and more in the direction of analytics, he was initially hesitant to sell off their insights into their users in a haphazard way. And so, for the first year, Zark had mostly turned down meetings with other companies seeking to monetize or capture this data.

But in 2017, Zark was approached by a company called PerPol, a data analytics and data mining company that had expressed keen interest in mining the sort of metrics used by *End*. The representatives of PerPol explained that they had begun a bold new initiative meant to use the sort of data collected by *End*, and the personality profiles they constructed from such data, to better understand and address the growing crisis of mental-health issues among young people – specifically, they were focusing on the district of Metavo, Texas, where their corporate headquarters were located.

They had amassed data that suggested that the mental-health epidemic in Metavo was correlated with another grave problem this district was facing: a spike in dropout rates. PerPol representatives told Zark all about a particular local school, Turing High School, which had just recorded its highest dropout rate ever in 2017, and where an alarming amount of students had been reporting mental-health issues in recent years. PerPol had set their sights on running a pilot program at Turing High.

To their surprise, when PerPol first approached the school board at Turing High, they were very receptive. The school board and administrators at Turing High had already been struggling internally over how to address the problem, and they had just been informed that the school risked being closed down entirely if they didn't reverse course in the next academic year. In the last two years, teachers and counselors at Turing High had been reporting alarming rates of mental-health related issues among their students, and apparently related metrics (like drop-out rates) had shown similarly dismal numbers. In their regular school meetings, teachers had arrived at the point of expressing despair at the numerous past failed attempts to address this suite of issues. The board at Turing High thus saw a great opportunity with PerPol and were eager to hear them out.

In their meetings with Zark, PerPol expressed interest in working closely with *End* on their pilot program at Turing High, where they would designate time every school day where students would play a video game designed by *End*, and where data from gameplay would be collected and analyzed by tools developed by *End* and PerPol. Zark was sufficiently intrigued by this initiative, and had two kids of his own (whose mental health and academic performance he sometimes worried about), so he decided to meet with school board members and administrators and Turing High.

In their first meeting as a group, Zark spoke convincingly of the promise of the novel AI approaches deployed by *End* and how they had reason to believe they might use the data collected in-game, together with other data collected by the school, to provide unparalleled insights into the mental-health crisis and growing dropout rates that had been troubling Turing High. Anxious to try something new, and emboldened by the ambition of Zark, the school board was eager to collaborate. They offered to Zark other data they had already collected on students: grades, attendance records, disciplinary history. As Zark observed, there was other data the school could provide as well: for instance, the school's internet network already monitored the internet use of its students and, through their phones, tracked their location while on the school's premises. Persuaded that this additional external data would better enable *End* and PerPol to isolate the causes of, and ultimately

help reverse, the trend of increasing mental-health issues and drop-out rates, the school board agreed to supply that data as well. They all agreed that understanding what causes students to drop out or first report mental-health issues may help them determine the right interventions and reverse course.

After some meetings, together they formulated the following goals of the pilot program dubbed “Kombat Dropout”: (1) *To identify predictors of student mental-health issues* as an indicator for dropping out, and then *to apply AI tools to these predictors* to alert the school to at-risk students; (2) *To equip teachers with tailored information and recommendations to assist students deemed “at risk”* by suggesting specific interventions based on their profile assessment.

In the subsequent months, the school board at Turing High agreed to let the pilot program proceed, and to provide the PerPol and *End* system with any student data collected in the past and present. Students and parents at Turing High were only told that they were now going to be a part of a cutting-edge “ambitious educational initiative” that would require that students re-allocate up to one hour each school day for playing “educational video games.” Beyond that, they were not informed of the specific agreement, and students were not given the opportunity to opt out (or given reason to believe they might want to). The administration and board members at Turing High knew how tricky it could be to get parents and other concerned parties to agree to any new initiative, let alone one this experimental and ambitious in design. Moreover, Zuck and others had said that they have reason to believe that “if a player believes they are being ‘analyzed’ in game-play, they will adjust their behavior and choices, skewing the models, which could detract from the predictive power of their analytics.” They accordingly argued that keeping parents and students in the dark was justified, on the grounds that their position required of them that they, above all else, promote positive educational outcomes for their students – this being a unique opportunity to fulfill this mandate.

In September 2018, “Kombat Dropout” began at Turing High, and the data started coming in. They started out by collecting any and all game data of possible relevance, but soon began looking at a large number of potential predictors, ranging from various in-game metrics to external factors (such as academic factors like grades, history of disciplinary action, attendance records, but also location-tracking data of students while on campus). *End* and PerPol harvested data for the entire academic year, from September 2018 through June 2019, and they were able to correlate any data obtained during that period with data supplied by the school on mental-health and dropouts over the

last 25 years. *End* and PerPol were able to generate powerful inferences that, they determined, would not have been possible if they had not had access to the non-game-play data supplied by Turing High.

Ultimately, the system was able to identify a number of robust “personality profiles,” on which were based 10 key indicators that, together, predicted whether a student would drop out with 90 percent accuracy, predicted whether a student would report mental-health issues for the first time with a striking 96 percent accuracy, and even gave insight into how mental-health factors affected risk of drop-out. As promised, they used these profiles and predictors to supply teachers with the relevant information from the profiles of at-risk students that both helped them understand why an individual was struggling and suggested targeted treatment approaches for improving their situation.

The majority of teachers at Turing High followed the recommendations made to them, and very quickly teachers were reporting greater attendance, classroom engagement, and improved testing scores. Moreover, there were numerous indications that there were improvements in the students’ mental-health (further corroborated by the independent reports of school counselors and surveys conducted by specialists). The administration, for their part, used information provided to them by *End* and PerPol to adjust certain aspects of the school environment in order to nudge students in ways deemed conducive to the improvement of academic performance and mental-health (for example, more resources were allocated to certain after-school groups, schedules were adjusted, etc.).

By the end of the academic year, Turing High seemed to have reversed course, and greatly improved its situation. Based on survey evaluations and other data, the mental-health of Turing students had seen substantial improvements since “Kombat Dropout” began. Furthermore, the drop-out rate fell from over 11 percent to just 4 percent – which was lower than the average for their district. And in 2019, more students than ever in Turing High’s history were accepted to four-year colleges.

In their latest community meeting, the administrators at Turing High revealed to the broader community all the work that had been done by *End* and PerPol, and credited them with the marked improvements. However, the apparent success of “Kombat Dropout” was somewhat unsettled by concerns raised by students and parents when they were finally informed about the full scope of the program. Students and parents learned for the first time how *End* and PerPol, with the full cooperation of Turing High’s school board and admins, had been using extensive student data (from game-play, but also external data) to make its recommendations. They were also informed that this

pilot program was the first of many envisioned by *End* and PerPol – who were now entertaining bigger ambitions – and that the data collection would continue at Turing High indefinitely, in order to track its success, and reinforce the newly attained positive outcomes.

While many parents were indeed impressed by the swiftness with which Turing High was able to reverse course; and while many students reported enjoying the game-play itself; the decision to run the pilot program, together with the decision to make it ongoing in the indefinite future, caused concern among many students, parents, and teachers. As some teachers and parents began to publicly voice their worries and criticisms, the story got wider attention. Eventually, the wider community of gamers got involved and were particularly vocal in their worries. Many expressed concerns that a player's behavior in-game could affect how you're treated in the real world. One such concerned citizen said:

What happens if game data is used in other contexts, or by other agents, so that in 5 years you get rejected at a job interview because a game you once played revealed you're not a team player? What happens if you can't get a reasonable insurance quote because game-play reveals you're especially vulnerable to addiction? What if you cannot get a loan for a house because a game you played revealed you were especially prone to "risky behavior"? What if this data is fed to predatory advertisers and application developers who weaponize that addiction, designing platforms, games, and content that use the data to get better and better at manipulating your physiology and dopamine responses?

Privacy experts soon weighed in and expressed further concerns about how information collected in games could be connected to a patchwork of online services and then used in a myriad of exploitative or detrimental ways. As many noted, single or isolated data points – such as decisions made within a game – may appear harmless when fed only to game developers; but, combined with other data or fed to other actors, such data harvesting can yield powerful tools for controlling behavior and psychology – with all the potential for abuse this implies.

Exercise Questions

1. **Stakeholders:** Who are the relevant stakeholders for the “Kombat Dropout” program?

List at least 5 groups of stakeholders, and rank them based on the relative degree of importance you believe ought to be attached to effects on them and to their concerns. Briefly defend your ranking. For instance, while

concerned parents might be regarded as relevant stakeholders, one might argue that effects on the students themselves ought to be given greater weight than the concerns of, or effects on, their parents – since the students are the ones more directly affected by the system, and more weight ought to attach to agents directly affected, as opposed to indirectly affected.

2. Benefits: *This question has 2 parts.*

- a. List 5 of the most significant benefits of the use of “Kombat Dropout” at Turing High?
- b. For each of the 5 benefits identified in part (a), present an alternative path towards the same end.

3. Potential Abuses/Misuses: *This question has 2 parts and asks you to consider some potential harms/abuses/misuses in using the predictive tools developed by End and PerPol in other similar contexts.*

- a. Suppose you are in Zark’s position – and, seeing the success of “Kombat Dropout,” you are now emboldened to take your tools to other applications. In this part of the question, you are to consider potential abuses/misuses for such a system. In particular, be guided by the “Thinking about Bad Actors” flowchart process; you will use this process to help identify potential ethical risks of such a program, when implemented in other settings.

First, consider at least 2 “bad actors.”

Then, following the flowchart, for each of the 2 bad actors:

identify a list of potential misuses/abuses by such an actor; then

identify a list of incentives/openings that may allow for or encourage such misuse/abuse; and then

identify a list of viable alternative procedures that might accomplish the same aim/benefit while removing/replacing the incentives/openings that allow for the misuse.

- b. Following up on the previous question, for each incentive/opening to each potential abuse/misuse associated to each of the 2 bad actors,

propose and defend what you think is the *best* of the alternative measures that could be put in place (where “best” should be understood as that path that most reduces the likelihood of the abuse/misuse in question, while continuing to preserve the likelihood of achieving the desired positive outcome).²

² As an example, focusing on the present case, one might raise concerns over privacy violations and abuses that could follow from such violations, discussing agents that could abuse such features, and identifying the specific features of the system that allow for such abuse. You might then have considered some alternatives. For instance, first, the databases could be pseudonymized and no record kept of the link between data points and students’ identities in the raw dataset. Second, even if the data is instead kept identifiable, you might have proposed that meaningful consent be gotten from students or their parents, who would also be shown

4. Who is in the Room: *This question has 2 parts.*

a. As described in the above case, Zark, the executives at PerPol, and members of the school board and administration at Turing High were effectively the sole decision-makers in deciding to move forward with the initiative. Who – of the relevant stakeholders you mentioned in your answer to Question #1 – should have been involved in the decision; and how/to what extent?

b. Now consider a group less likely to be included in the decision process – one that you did *not* originally include in your list of stakeholders – but that will still be affected by the decisions. What are some concerns such a group might have? Name the group and then list at least 2 concerns they might have.

5. Value Conflicts and Trade-offs: In general, when designing a computerized system (especially one of AI governance, used to guide decisions involving elements of behavioral “nudging”), some trade-offs are inevitable. For example, individual privacy considerations may need to be balanced against the desire to achieve legitimate social ends. The extent to which specific values are embedded in systems reflects the priorities and preferences of the systems’ designers.

This question has 3 parts.

a. Discuss 1 important conflict of values (other than that between data privacy and the effective achievement of certain social changes) presented by the above case.

b. How should decisions about the appropriate balance between these two values be made?

c. If a compromise/trade-off has to be made, requiring that one of the two values is sacrificed or minimized in some way, is this justifiable – and on what grounds? Working with the conflict of values isolated in part (a), discuss a similar situation (relevant to the above case) where a potential sacrifice would need to be made, and present an argument for such a trade-off.

6. General Ethics (Consequentialism vs. Non-Consequentialism):

Some ethicists argue that certain actions are impermissible regardless of what (unambiguously) positive outcomes they might bring about; others believe that the ends may justify the means, especially if the “total

the arguments for the actual plans for use of the data. Then, you might present an argument that, for instance, the first alternative is the best of the two, since, in this case, informing the students of the uses of the data might reduce the reliability of the data collected. On the other hand, you might argue that in an educational setting, students ought not to be treated as research subjects, especially if they are not informed that they are being used in this way.

effect” (on most relevant stakeholders, on most reasonable timescales) is net positive. In the above case, the school administrators, Zark, and those at PerPol, seemed to have both good ends and – at least as they saw things – appropriate means.

In computerized systems (and AI systems in particular), it may be extremely challenging to even keep track of the various means in use, and determine an appropriate hierarchy for each of the steps involved. If all the means must be evaluated independently of the ends, or some ultimate end, that they generally bring out, it may be very difficult or impractical to evaluate the permissibility of different actions. Considering this difficulty of assessing each of these steps in their entirety, school officials and Zark would likely argue that it was most reasonable to focus on achieving the noble end of reducing the student dropout rate.

Yet even if nearly everyone felt the school dropout rate was a problem; and even if the steps taken to address this problem indeed altogether amounted to a viable solution to that problem; it is likely that not all stakeholders would agree with Zark, or the school administration, about the appropriateness of their means (in particular, their use of student data without consent to produce unauditible results, and without providing students with their risk profiles). They might argue that the way Zark and the administration went about reducing the dropout rate, and the violations they had to effect to do so, undermined the legitimacy of its “noble” end.

What do you think? Do the ends justify the means in this case? How about in general? Or do you believe there are certain things that just must not be done (regardless of how good the consequences are)? However you respond, be sure to defend your answer.