# The Ethics of AI in Games

David Melhart ⬤, Julian Togelius, Benedikte Mikkelsen, Christoffer Holmgård,
and Georgios N. Yannakakis ⬤, *Senior Member, IEEE*

*Abstract*—**Video games are one of the richest and most popular forms of human-computer interaction and, hence, their role is critical for our understanding of human behaviour and affect at a large scale. As artificial intelligence (AI) tools are gradually adopted by the game industry a series of ethical concerns arise. Such concerns, however, have so far not been extensively discussed in a video game context. Motivated by the lack of a comprehensive review on the ethics of AI as applied to games, we survey the current state of the art in this area and discuss ethical considerations of these systems from the holistic perspective of the *affective loop*. Through the components of this loop, we study the ethical challenges that AI faces in video game development. *Elicitation* highlights the ethical boundaries of artificially induced emotions; *sensing* showcases the trade-off between privacy and safe gaming spaces; and *detection*, as utilised during in-game *adaptation*, poses challenges to transparency and ownership. This paper calls for an open dialogue and action for the games of today and the virtual spaces of the future. By setting an appropriate framework we aim to protect users and to guide developers towards safer and better experiences for their customers.**

*Index Terms*—**Artificial intelligence, ethics, video games, affective computing.**

## I. INTRODUCTION

**V**IDEO games are key to our understanding of human behaviour due to their vast popularity, the multi-modal ways players can interact with them, and the various ways games can express emotion and adapt to a player's style. Even though values such as transparency, trustworthiness and responsibility are core aspects of ethical systems in other domains, video games present unique challenges in terms of ethics. Dark patterns in game design [1], predatory monetisation strategies [2], and the black-box nature of games hinder transparency [3] and raise several ethical concerns. These issues are far-reaching from game design and development [4], [5] to societal impact and research ethics [6].

In this survey paper, we aim to address the ethical considerations of game AI tools and methods through the lens of *affective computing*. In particular, we focus primarily on *player modelling* [7] as a field of game research that considers the aggregation, simulation [8], and understanding of gameplay and user experience in games. We, thus, structure the discussion
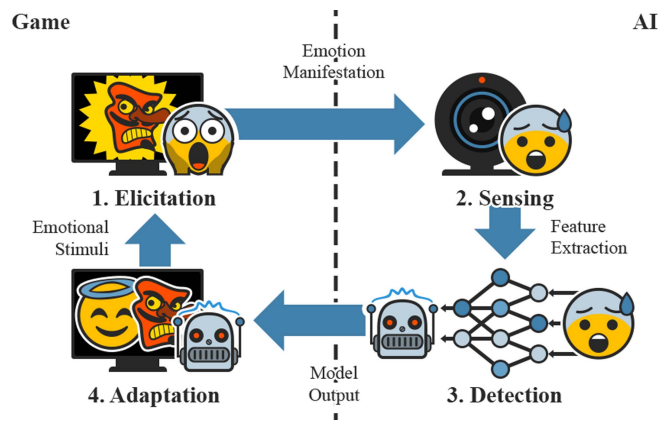
Fig. 1. *Affective Game Loop* [9]. The loop relies on the game's parameter space to *elicit* an emotional response. This response is *sensed* by an AI model that *detects* change(s) in the player's emotional state. The output of the affect model can be used to *adapt* the game content and generate a new set of stimuli for the player.

of AI ethics in games around the *affective game loop* [9] (see Fig. 1). The affective game loop describes the relationships between emotion expression, elicitation, detection, prediction, and subsequent reaction. It presents a complex game system which facilitates these processes and adapts to the user's emotional response. This loop can assist AI systems to generate personalised aspects of games such as agent behaviour, levels and images [10] or guide an orchestration process [11], [12] across creative facets such as text, levels and visuals. The concept of the affective game loop has been explored thoroughly in academia [13], [14], [15], [16], [17]. Meanwhile, the adoption of affect-driven adaptation systems in games has been gradual over the last twenty years; indicative yet representative examples include *Façade* (Procedural Arts, 2005)—see Fig. 2—and *Nevermind* (Flying Mollusk, 2016)—see Fig. 3.

The paper is structured as follows. After an overview of related literature (Section II), we discuss the ethical dimensions of game AI through the phases of the *affective game loop*—for a detailed structure of our survey see Table I. In particular, Section III covers aspects of *elicitation* and how dark patterns are used to manipulate and reduce the players' emotional agency in harmful or exploitative ways; Section IV takes a thorough look at *sensing* and issues related to the tradeoff between privacy and control, and malicious action in games; Section V discusses affect *detection* and the complexities of transparency in limited information systems such as games; and finally Section VI reflects on questions of data and model ownership during the affect-driven *adaptation* phase. The paper ends with a discussion

TABLE I
ASPECTS OF THE AFFECTIVE GAME LOOP WITH THEIR ASSOCIATED AI VIRTUES (INTRODUCED BY BOSTROM AND YUDKOWSKY [18]); MAJOR PITFALLS; AND
POSITIVE INITIATIVES UPHOLDING AI VIRTUES AND BENEFITING END-USERS

| Affective Loop | Section | AI Virtues [18] | Major Pitfalls | Positive Initiatives |
|---|---|---|---|---|
| **Elicitation** | Section III | Responsibility, Auditability | Predatory monetization, dark design patterns [1], and generating harmful content | Governments tightening regulation around predatory monetization practices [19]; Prolific designers taking a critical stance against dark patterns (e.g. Six to Start CEO Adrian Hon [20]) |
| **Sensing** | Section IV | Transparency, Incorruptibility | Lack of transparency in data inferred by AI systems | Ubisoft's ML models sensing for toxic behaviour in *For Honor* (2017) [21] |
| **Detection** | Section V | Transparency, Auditability, Predictability | AI systems overfitting to skewed populations and perpetuating harmful historical biases | Game studios sharing datasets with the academic community (e.g. EA and Nintendo [22], Ubisoft [23] and Riot Games [24]). |
| **Adaptation** | Section VI | Responsibility, Transparency, Incorruptibility | Unclear chain of responisibiliy and ownership of data and output of human-AI co-creation | Microsoft's *Xbox Transparency Reports* showing actions taken in content moderation[1]. |



Fig. 2.   In *Façade* (Procedural Arts, 2005), the player can interact with the game's agents through free-form text. The underlying AI responds to the player input based on its semantic and emotional content.
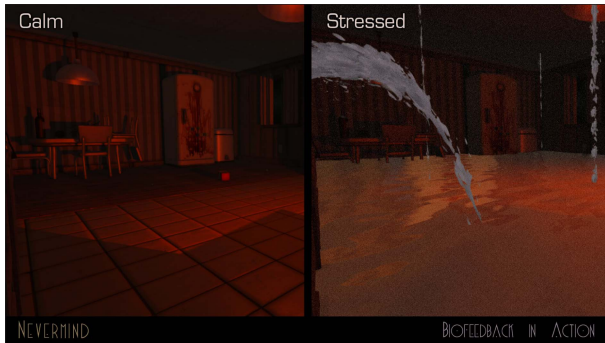


Fig. 3.   In *Nevermind* (Flying Mollusk, 2016), the player explores dream-like horror environments. The game content is adjusted based on the player's emotional state by introducing more dangers as the player's stress level increases.

on several other issues related to game AI ethics including AI algorithmic biases, compute fairness, and in-game toxicity and violence.

## II. RELATED WORK

This section reviews the literature on ethics research and ethical frameworks in AI and games research.

### A. Ethics in Artificial Intelligence

Ethics has been a constant challenge in the field of AI fuelled by academic and practical interest into the governance of autonomous systems and public anxiety towards data-driven black-box infrastructures [25], [26]. Ethical frameworks have been developed to address these anxieties, generally aiming to provide guidelines for creating beneficial, transparent, and trustworthy applications.

One of the most popular ethics frameworks applied in AI consists of the virtues of *Responsibility*, *Transparency*, *Auditability*, *Incorruptibility*, and *Predictability* introduced by Bostrom and Yudkowsky [18]. The *responsibility* of AI algorithms refers to their clear oversight on the chain of responsibility as the output of the algorithm can be attributed to either individuals or organisations [27]. As we will discuss in Sections III and VI, a clear chain of responsibility is often lost between the original data, the inferred models, and large-scale ensemble architectures using third-party AI.

*Transparency* is one of the more complex ethical virtues and cornerstones of AI Trustworthiness [28], [29]. On the one hand, it can refer to a kind of *algorithmic transparency* that promotes AI decision-making processes that are explainable [30] and clearly understood by their users [27]. On the other hand, it can refer to a *systemic transparency and openness* of AI-powered applications; that is legal access to AI infrastructures themselves [28]. As we will show in Sections IV, V, and VI, the industry has a troubled relationship with transparency, with many companies not disclosing their use of player models to gain further insights from user data. While transparency in relation to direct data-collection is clear, "inferred" information such as computational models and their output is much less protected by legal frameworks [31].

*Auditability* implies that the correctness of the output of AI systems should be verifiable by a third party. As we discuss in Sections III and V, auditability and general transparency is a serious blind spot of the video games industry. Although some of this blind-spot can be attributed to the inherent opacity of AI systems [32], there is a definite limitation raised by legal opacity restricting access to AI architecture and training data by external auditors [33].

*Incorruptibility* means that the system is robust against manipulation. Even though the obfuscation of datasets, algorithms,

and their output definitely provides some level of protection, obfuscation is fundamentally clashing with the principle of transparency. Due to their interactivity games are under constant siege by malicious users, however, their corruption is not necessarily an outside force. As Gebru points out, AI bias tends to exacerbate the sociopolitical and socioeconomic disparities in our society as they perpetuate inherent biases of the creators of AI models and our social reality [34]. As we discuss in Section IV, while one of the primary goals of applied AI in the game industry is to increase the robustness of systems against external attacks, there is much less discussion and transparency about the inherent bias in the employed AI systems.

Finally, *predictability* refers to self-consistent AI outputs and algorithmic behaviour. Predictability is a less prominent yet important aspect of AI ethics, which aims to push applications towards a more reliable and fair implementation [27]. Predictability goes a long way towards eliminating AI bias, which we detail in Section V. The aforementioned virtues are being understood as the cornerstones of AI ethics and solidified [3], [35]—in some shape or form—in the *IEEE Ethically Aligned Design Guidelines* [36], the *Humane AI Ethical Framework* [27] and the newly emerging concept of *Trustworthy AI* [29], [37], which also plays a fundamental role in the new *Ethical Guidelines for Artificial Intelligence of the European Union* [38].

A recent meta-review by Yu et al. [26] of the AAAI, AAMAS, ECAI and IJCAI conferences mapped out the field of Ethics in AI (EAI) and identified four major areas under this domain. The first category is research focusing on leveraging AI techniques to *explore questions of ethics* faced by humans. The second and third categories focus on internal *decision-making frameworks* for AI agents acting either as individual units or collectives. Finally, the last category focuses on *ethics in human-computer interactions*. For a complete review of all these avenues of research we refer to Yu et al. [26]; here we focus only on the latter category as it is the most relevant to the domain and purposes investigated in this paper. As positioned by Yu et al. [26], [39] and echoed by the larger research [3], [35], [40] and policy making [36], [37] communities, ethical HCI systems should conserve the autonomy of humans, be beneficial to the user, and minimise underlying risks. Summarising this sentiment in relation to affective computing, the *IEEE Ethically Aligned Design Guidelines* explicitly state:

> "To ensure that intelligent technical systems will be used to help humanity to the greatest extent possible in all contexts, autonomous and intelligent systems that participate in or facilitate human society should not cause harm by either amplifying or dampening human emotional experience" [36, page 6].

Beyond the scope of emotional autonomy, however, there is also the question of transparency and autonomy in human-AI interaction in general. Rovatsos raises the issue of the general distrust towards machines and whether it is ethical for an AI system to conceal itself [41]. Although it can be easy to consider total transparency as the most ethical, the issue is more complex. A new ethical conundrum emerges when we consider that in some human-computer interactions, a lack of transparency can improve the efficacy of the system [42]. If this

is true, wouldn't the performance drop—that was induced by increased transparency—hurt the user in the long run? Would in this situation total transparency take away from the user's autonomy? On the other hand, could an opaque system even present fair choices to the user? These questions—raised by Rovatsos [41]—presuppose a benevolent system. However, AI is not always designed to be benevolent. Perhaps the most striking example of this is lethal autonomous weapon systems, which are designed to kill humans without considerable oversight [43]. Even though real-life killing robots might seem to be removed from the domain of games, pushing a military agenda and aiding both recruitment and research has never been far from video games [6]. And as we discuss at many points in this paper neither is emotional exploitation nor psychological manipulation.

AI researchers from the fields of computer science, engineering, robotics, medicine, games, and more are calling for stronger regulations on exploitative and harmful AI and a push for benevolent AI applications [6], [43], [44]. Their fears are not unfounded as current research and industrial application of AI are more than capable of exploiting and harming humans en masse, from social engineering [45], through psychological manipulation [2] and exacerbating existing socioeconomic disparities to physical harm [6], [43].

### B. AI Ethics in Game Research

AI ethics in game research is a fairly under-researched area. The handful of papers existent in the literature focus mainly on player modelling [3], [46], ethical development practices [4], [5] and ethical practices in research [6]. In contrast, more work has been carried out on games as ethical systems [47] and the outcome of responsibility of game design [1].

In a review of the field of player modelling, Mikkelsen et al. provided an overview of emerging ethical issues [3]. In their analysis relying on the framework laid out by [18], Mikkelsen et al. identified a number of areas of concern from monetisation through content management to dynamic adaptation and privacy. Most issues emerging in these areas are connected to the lack of *transparency* and *auditability* of computer models, especially in industrial settings. One solution to the lack of transparency and interpretability is offered by the field of *explainable AI* [48], [49]. A possible avenue for adapting explainable AI is through *open player models* [46], which are based on Open Learner Models applied to games [50]. Open player models incorporate an explanatory module into a given AI application which gives clear feedback to the user on the behaviour and predictions of the model. As the module providing transparency is removed from the main pipeline of the algorithm, in principle open player models can be cost-effective to implement in existing systems as well. Nevertheless, although explainable AI principles can help build more transparent systems in theory, the practical application of such frameworks appears to be challenging. Black-box algorithms such as deep learning neural networks are very popular in data science due to their performance, and they are notoriously hard to explain and interpret—despite advances [49]. On the other hand, effective white-box systems are still an open challenge to the field [48] and might not even be

sustainable or desirable from a business perspective as the games industry is known to treat datasets, data-processing pipelines, and AI models as strictly-kept trade secrets.

Beyond the concerns of transparency there is an alarming issue of intentionally harmful usage of AI models that exploit addiction and irresponsible spending habits [3], [5]. Despite a growing concern against aggressive and deceptive monetisation techniques—often aimed at children—there is still a lack of legal and practical frameworks that are capable of addressing such issues [2]. King et al. [2], for instance, examined 13 different patents connected to video game monetisation and found that almost all of them relied on the exploitation of the players' data to optimise the delivery and timing of ads and purchase offers. They note that with the expansion of AI methods it is expected that such systems will become more sophisticated and ubiquitous in the future, making the issue of ethics in player modelling more pressing than ever.

### C. State of AI Ethics in Practice

Although ethical frameworks have been developed to provide guidance, the lack of specificity often leads to a small scale of adoption. If we look at the core issue of *transparency*, which is also often required for the assessment of other components of AI trustworthiness, we find that both affective computing and games applications are lagging behind [51]. This is true despite the issue of transparency being propped up by more robust legal frameworks than many other components of ethical AI. In the European Union, the General Data Protection Regulation (GDPR) [52] is meant to give a legal framework and transparency to data handling (also in an AI context). However, a review of *serious games*—games developed for healthcare, educational, hiring, or other non-entertainment purposes—found that two years after the adoption of GDPR, it has had little to no effect on the research community [53]. Similarly, in a recent exploration of affective computing through the lens of GDPR laws Hauselmann found that the field faces serious issues in terms of *transparency*, *responsibility*, and *predictability* [54]. Hauselmann highlights the delicate nature of emotional data as something that is not necessarily protected under current legal frameworks but extremely personal to the users. However, the question of emotional data is further complicated by the fact that while user behaviour is relatively easy to observe and record, emotional data is often extracted through means of peripheral signals and machine learning. In this sense affective data is inferred and not observed [31]; as a result, the majority of affective computing applications appear to be inherently opaque. As there should be a right to an accurate portrayal of personal data, inaccurate predictors might infringe on the personal rights of users. This is hard to prove, however, as these models are often difficult to audit. This phenomenon is amplified because there are fewer practical concerns for inaccurate models up to a certain degree. Often even if a user is profiled inaccurately, an imperfect prediction can still be used to great effect in an adaptive system [55]. Moreover, commercial applications often safeguard their models as trade secrets or cannot handle the constraints and overhead of implementing ethical safeguards on a fundamental level.

The above examples focus predominantly on the research community; in the games industry, the problem of transparency can be even more prevalent. More often than not, users are unaware of the data collected and inferred by algorithms. As Kröger points out, data collection in games is generally made invisible to the players as it is "woven into a game's environment" [51]. Given this opaqueness and a blasé attitude of users towards—what they perceive as—anonymous play, it is questionable to which extent regulations such as the aforementioned GDPR could reasonably be upheld. A thorough review of five companies by Vakkuri et al. [35] revealed that even though developers might consider ethics as an important question, they have little to no tools to address it in a systematic manner. Mitigation of ethical risks in AI systems thus becomes low-priority and generally addressed in a post- and ad-hoc manner, if at all. Reviews of game industry applications that span from player modelling [3], [46], to data-driven game development [5], and to procedural content generation [4] reveal a similar pattern. Because the games industry is a fast-moving field with growing pressure on producing more and more content with the advent of live-service games, the application of ethical frameworks to AI in games—including but not limited to player modelling—remains an after-thought without clear ways to integrate the mitigation of ethical problems into existing industry pipelines.

## III. ELICITATION – BOUNDARIES OF ARTIFICIALLY INDUCED EMOTIONS

We start examining the affective game loop from the *Elicitation* phase. Doing so we are faced with the ethical boundaries of artificially induced emotions. Although inherently personal and subjective, emotions do not enjoy legal protection to the same extent as other personal data [54]. The core issues we encounter in this area are ownership and autonomy over one's own emotions. The so-called *dark design patterns* [1] have been used in games to compel players' behaviour through affective manipulation and with the advent of Big Data analysis and machine learning, there is a potential for a new wave of dark design patterns [2]. As games are often marketed towards children, the ethical side of the emotion elicitation in games, their use and their goal have to be considered. Importantly, the challenge of dark design patterns is core to game design principles but not necessarily to the AI algorithm associated with a game. One should thus take a dive into the problematic ethical aspects of the game design prior to examining the role of AI within a particular game.

While a few years ago *loot boxes*—i.e., virtual items that can be redeemed for other random items that provide some value to the players [56]—made waves [57], [58], [59], the new monetisation technique sweeping across the industry is the *battle pass* or *season pass* system. Unlike previous iterations of *premium subscriptions*, *in-game currencies*, *downloadable content* packs, *loot boxes*, and *gated progression*, the *battle pass* system does not promise any immediate tangible reward to players. Instead, players buy into access to time-limited content

updates, which they still have to unlock in-game within a given time frame [60]. This type of monetisation reformulates the value proposition of online games and shifts the focus from commodities to services [61]. While the loot boxes of yesteryear were designed to operate on the same psychological buttons as gambling [57], [58], emerging battle pass systems build more on a *feeling of missing out* [60] and societal pressure [62]. In many modern online games—such as *Fortnite* (Epic Games, 2017), *Apex Legends* (Respawn Entertainment, 2019), *Fall Guys* (Mediatonic, 2020) and *Overwatch 2* (Blizzard Entertainment, 2022)—these monetisation practices often coalesce into a virtual storefront, where in-game currencies can be bought for real money, then spent on single purchase upgrades and battle passes alike.

Where player modelling techniques can make loot boxes, battle passes, and other similar monetisation techniques more concerning is the ability to target people more prone to spending. Predictive models have already been in place for years in the industry for the estimation of churn [63] to keep track of players lost and their velocity through a game. Similar models, however, can also be used to find and target potential excessive spenders—often called "whales" in the industry [64]. Affective computing models estimating the users' emotional state can be used for the targeted and timed delivery of ads and promotional offers to maximise user spending. Affective modelling methods, however, can also help build more responsible systems that detect a risk of a problematic behaviour. Out of the psychosocial aspects of addiction (salience, mood modification, tolerance, withdrawal symptoms, relapse, conflict) [65], affective computing methods could be especially useful to pick up on mood modification and tolerance (as a diminishing affective response) at the very least and flag users as at-risk consumers. In the same way, machine learning models based on affective and behavioural feedback [66] are deployed to target monetisation and retain consumers, they can also be used to deploy "precision psychiatry" [67]. Most notably, EA was accused to leverage their patented dynamic difficulty adjustment system to push players to spend more money on loot boxes [68]. Even though the case was dismissed and EA swore to uphold "fair play" in their online games [69], this indicative example goes to show how algorithms can be used to elicit emotions that influence players to act against their own interest.

While affective computing systems in general often obscure how they infer information and predictions [54], the black-box nature of games is even more apparent. Games are often viewed as "smoke and mirrors" when it comes to the dichotomy between the game's parameter space and the conveyed aesthetic. Game designers are invested in hiding what lies within the game rules to facilitate a suspension of disbelief and thus make the experience more impactful and believable for players [70]. Often game designers are relying on transgressive aesthetics [71] to create experiences that have a larger emotional weight. Although the area is still largely unexplored some games like Flying Mollusk's *Nevermind* (2016) integrated computer models to guide their elicitation. In *Nevermind* the game's transgressive aesthetics is amplified when the player is under stress and subdued when they calm down. Intentionally transgressive content—even when

controlled by autonomous systems—can be tuned and managed more consciously; unintentionally harmful or offensive content, however, poses a much more complex problem.

Of course some more obvious errors are easier to catch with simple pruning, but computational models can also encode biases—like gender biases or offensive stereotypes—which in turn result in harmful content [72], [73] unintended by the designer. A good encapsulation of the complicated nature of using AI generators, content moderation, and privacy is *AI Dungeon* (Latitude, 2019), a text adventure game based on the GPT-2 [74] and GPT-3 [75] language models. Latitude has become the focus of a controversy when they decided to take action against offensive content *AI Dungeon*—mainly involving stories containing non-consensual sexual content and child pornography [76]. Although much of the questionable content banned by Latitude was generated deliberately by their users, the model was also known to generate sexually explicit content seemingly unprompted, including "writing children into sexual scenarios" [76]. Users raised concerns about the decision of Latitude to address the issue with strict moderation, automatic flagging of problematic materials, and monitoring the content of users' privately generated stories. On one hand, this decision pushed all the *responsibility* of the content to the users even though it was co-created with Latitude's algorithm; on the other hand, the human moderation of private content raised privacy concerns. The controversy showed that despite the best intentions of Latitude, a lack of *transparency* and a clean line of *responsibility* [18] leads to detrimental outcomes for both the company and the end-users. The swift shift in how moderation was done on the platform made the already black-box system even harder to navigate for players, which in turn reduced both the transparency and the users' trust in the system. Many players felt unfairly flagged for content that either fell within community guidelines or was generated by the model virtually unprompted [76]. As Latitude was not the developer of the underlying foundation language model [77], it was unclear how the system can be effectively *audited* and since there was no established *responsibility* for the system, all the blame fell to users who interacted with the model.

As we can see elicitation through AI-assisted systems has potentially harmful effects on the end user. On the one hand, video game companies can rely on affective computing models to fine-tune and personalise targeted monetisation strategies. This carries the danger of intentionally or unintentionally facilitating addiction or pressuring users on an emotional or social basis. On the other hand, generative systems can be unreliable and surprising and generate unwanted content without the designer's knowledge or the user's consent. To prevent subsequent issues, generative systems in games should demarcate a chain of responsibility for the model's output and offer tools to players to mitigate unwanted content. Even though foundation models offer a robust solution for generating the content, often the complexity, black-box nature, and ownership of the models limit the *auditability* of these algorithms [77].

While leveraging predatory monetisation strategies has been a prevalent pattern—especially in the mobile games industry—not all studios have followed suit. A good counter-example is Six

to Start, the developer of the immensely popular *Zombies, Run!* (Six to Start) (2012) mobile exergame. The studio has not only forgone the usual dark design patterns, showing that games can be successful without putting psychological pressure on their players, but they have also been making a firm public stance against these practices [20].

## IV. SENSING – PRIVACY AND CONTROL

Following elicitation, the next step of the affective loop is *sensing:* the capture and processing of the manifested emotions. The central issue of sensing is that of privacy—i.e., when, how, and kind of data is being captured. While user privacy might define a clear issue in other domains, games present a special case. Particularly due to the interactivity of the medium, a certain amount of dynamic control is needed to maintain oversight over toxic and malicious actors in a game's ecosystem and prevent unintended or negative effects on the players' mental well-being.

It has been shown that users of affective computing systems prefer clear notifications and potential control over the output of sensors [78]. This need is also supported by the *principle of autonomy* when it comes to ethical applications [79], and more concretely by the *right of accurate portrayal* [54]. Unfortunately, in multiple instances affective computing applications fail to address these needs as the data and data pipeline is kept secret from the user [54]; game companies are no exception to this. In a recent comprehensive study on data types collected by the video game industry, Kröger et al. [80] reveals that companies are collecting and inferring a wide range of data often without the user's knowledge.

Most of the data collection focuses on in-game behavioural metrics which can be used to predict player skill, preferences, or content consumption and spending habits. The inference of spending habits and personal identifiers—such as location, gender, financial status, etc—can clearly fuel harmful policies. Even game mechanics, which on the surface are there to benefit the player—such as matchmaking or dynamic difficulty adjustment—have led to concerns in the past. As mentioned in Section III, one of the most recent examples of dynamic difficulty adjustment working against the users was the case of EA Games. In a lawsuit, EA was accused to use difficulty adjustment to influence player spending on loot boxes [68]. The lawsuit itself was later dismissed [69], however, this case still highlights the public distrust towards systems that collect behavioural data. This is not to say that games rely solely on such data. As Kröger et al. point out, game companies are adapting sensor data in their datasets at an increasing rate [80]. Eye-tracking, voice data, GPS information, and peripheral signals from smart accessories can all be used to enrich game datasets and potentially reveal a wide variety of personal information about the user. In today's interconnected world, it is becoming exponentially easy to triangulate and infer the identity of players—using their in-game data, username, GPS location, preferences, and distinct play patterns they use to interact with the game [81], [82]—to the point where it is questionable whether game data can be truly anonymised at all [80].



Fig. 4. Traditional flagging methods for toxic behaviour in *For Honor* (Ubisoft, 2017). Such methods often depend on user reporting which, in turn, may allow many toxic events to remain unnoticed and stay unreported.

Complex player profiles can be built on inferred gender, age, socioeconomic status, and interests, which can fuel harmful models exacerbating problem behaviour such as gambling and excessive spending. Beyond static high-level profiles it is also possible to infer the emotional state of users through keystroke patterns [83], voice [84], or in-game behaviour [14], [85]. Although the secrecy of the industry is a major concern, there is a tradeoff in terms of privacy and autonomy that is afforded to the players. Similarly to how behavioural analytics is used to infer a rich player profile [86], emotional data can also be used to model and subsequently enhance the play experience [9], [87]. These types of tradeoffs were identified by Ishowo-Oloko et al. as the *transparency–efficiency tradeoff of human-machine cooperation* [42]. While not always applicable to human-computer interaction, there are instances where the inherent bias against AI [41] can hinder a human-computer system if total transparency is maintained.

Models incorporating emotional data can also be used to enhance other game systems involved in the moderation of user content and interaction. Most recently, Canossa et al. presented a robust method to flag the occurrence and severity of toxic and emotionally abusive behaviour in *For Honor* (Ubisoft, 2017) [21]. Although the input features of these models are mainly behavioural in nature, the inferred actions are emotional. While community guidelines are generally presented clearly, human moderation can become cumbersome with the breadth of data increasing with new players. Additionally, traditional reporting systems rely on user input, which could come with its own limitations including unreported events and subversion of the system by malicious users (see Fig. 4). As toxic players are often trying to find new ways to circumvent regulations, in the future automatic flagging systems—that incorporate emotional data as part of their input or output features—can make games a safer place for players. The aforementioned study is a good indicative example of how affective computing applications can be used to enhance the *predictability* of the system towards its users and how we can use AI to both deliver clear value to players and improve the game experience.

When it comes to privacy and *transparency* the state of practice in the game industry appears to be severe. The lack of consent, autonomy, or in many cases just knowledge about the collected data and its usage is a serious and prevalent ethical

issue. In many instances this ethical failing can be traced back to lax regulations, where legal requirements for consent and compliance are technically fulfilled but do not facilitate transparency in a tangible way. However, despite recent legal efforts to institutionalise privacy requirements such as the GDPR of the EU, many industry players are still falling behind, unable or unwilling to address the most pressing issues [84]. On the other hand, not all data is collected for the purpose of exploiting and manipulating users. Often a wide range of data—otherwise considered "non-essential"—can enable the automatic flagging and moderation of toxic and malicious users. Some players aim to actively harm others in- and outside of the game through emotional abuse. Affect-driven applications of AI can offer solutions for identifying both the occurrence and the impact of toxic behaviour, making online games a safer, more reliable and *predictable* space for everyone involved. For this to happen, however, clear rules have to be put in place and users must be notified of what kind of data is being collected for what purpose. Even though the detection of toxic behaviour and negative gameplay outcomes might be beneficial for the players, it doesn't mean that the principle of *transparency* cannot be upheld.

## V. Affect Detection – Transparency in Limited Information Systems

The third core step of the affective loop is *Affect Detection* which refers to the computational processing and prediction of certain aspects of affect [88]. One of the major ethical challenges of deploying affective models, in general, is their *transparency* towards users [3], [54]. More often than not companies are not disclosing that user data is modelled, let alone informing users about their system's predictions. Similarly to privacy challenges discussed in Section IV, the issue of transparency is of ambiguous nature too.

We have already touched upon the issue of games as limited information systems in Section III. Games often withhold information to create uncertainty, decrease cognitive load, and construct challenges for players [89], [90], [91]. Moreover, games limit the up-front information that players have access to so that they facilitate learning [90] and an experience of *flow* [92]. Abuhamdeh et al. have shown that greater outcome uncertainty does indeed lead to greater satisfaction when the player is succeeding [91]. They also found that as perceived competence rises, suspense and uncertainty become a major facilitator of *intrinsic motivation* in video games; for an overview of intrinsic motivation in games see [93]. Because uncertainty is a fundamental element of game design, it is very difficult to mitigate issues of *transparency*. While explainable AI frameworks generally advocate for open communication towards the user about model predictions [46], [48], when it comes to in-game adjustments this can be detrimental to the player experience [42]. It is important, however, that this tradeoff only applies to systems that use behaviour and emotion prediction to adjust in-game content, where the models only reach as far as the "magic circle" of the game experience [94], [95].

It is important to note that games do not exist in a vacuum and the experience is far from being a closed bubble. In systems where affect detection is used to inform monetisation strategies, users should be informed in a clear and comprehensive way about the output and the goal of the algorithm to preserve the system's *transparency* and *predictability*. Of course, as cited above, some game studios have a bad track record keeping users informed about their practices behind the curtains [84]. Although some of the secrecy can be chalked up to attempts to address the *incorruptibility* of the system by keeping it obscure, more often than not it seems video game companies rather want to protect their resources (e.g., in terms of trade secrets, trained models, and datasets).

Beyond the questions of *responsibility*, *transparency*, and *auditability* when it comes to detecting emotional and behavioural outcomes, game developers must also face the consequences of the inherent bias present in AI systems. For example, even though it might be responsible and beneficial to filter players based on certain emotional or behavioural states, models can also propagate unseen harmful biases. Even though *transparency* becomes a critical duty of developers in such instances, other ethical standards must be drawn as well to preserve the integrity of these applications and reconcile with the experience being provided to the players.

One of the most common causes attributed to algorithmic bias is a faulty dataset [96], [97], [98], [99]. The more apparent issues in this regard are a skewed population, lack of control for diversity, and the non-critical capture of historical biases [96], [97]. The latter of these issues makes it especially hard to mitigate algorithmic biases. On one hand, "clean", unbiased data might either not be available or impossible to attain. On the other hand, systems relying on historical biases can propagate patterns that seem true to a casual observer and are only revealed as biased under a more critical analysis [99], [100]. The issue becomes more severe due to the lack of *transparency* and *auditability* in the field. It is often next to impossible to recognise a biased dataset unless the algorithm breaches the trust of the users in a serious and very apparent way. The most prolific of these instances are tied to sexist and racist outcomes [34], [98]. In one instance *Google Photos'* algorithm was mislabelling pictures featuring black people as "gorillas" [101]. Of course, Google's algorithm was not created to be racist. The issue instead stems from the lack of diversity in the dataset that was used to train the model. As the model was trained on a dataset featuring predominantly white people, it learned to associate "whiteness" with "people". This error reveals a fundamental issue with a less-than-critical approach to data. Historic and institutionalised injustice defines our social reality. As injustice is ingrained at a fundamental level in our society, this type of bias is very hard to eliminate. The responsibility of the curators of large datasets and the developers of AI models is to apply critical forethought to processing and modelling to reduce the impact of these biases.

While models used in the video game industry can arguably fail in similar contexts, there are more potential pitfalls unique to games. Similarly to other skewed datasets, game data can also be skewed towards atypical players, deriving an unfair distribution of the population as a whole. One example of this would be the overabundance of data from players with large amounts of

playtime. Because they might not represent the population, an algorithm that focuses on these players would likely produce sub-optimal content for the remainder of the player base. This can especially be true in the initial phases of development as initial testers tend to be young and relatively good players. Another example of in-game bias would be discrimination towards atypical behaviour or emotional response. Systems monitoring toxic behaviour and bots are often based on high-level aggregated data on in-game actions and chat interactions [21]. If these algorithms do not account for diversity and expect a behavioural and emotional response based on a Western, neuro-typical audience, they might flag good-faith players who are not conforming to certain behaviour or communication standards. On the other end of the spectrum, outliers can outperform the expectations of the model and are labelled as cheaters as happened to Julias Jackson, an autistic boy on the *Xbox Live* ecosystem [102]. This error reveals the fragility of many automated systems when they have to apply their predictions outside of their trained boundaries. As we cannot be sure where those boundaries are in the wild [3], the challenge of organising and labelling our training data becomes a core issue. There is only so much we can anticipate in terms of future diversity requirements and even if we do, we might lack the tools to label our data correctly. Although it is relatively easy to rely on user reporting—especially when it comes to moderation—this also injects a large amount of bias into our systems both by malicious and good-faith users. Unfortunately, user reporting can often exacerbate other underlying biases, such as men accusing women of cheating for outperforming their peers [103]. Even though these AI tools are in most cases used as flagging systems and not automated banning systems, the story of Julias Jackson shows that the infrastructure is far from being perfect, even with human oversight.

Nevertheless, many industry leaders are well-invested in creating fulfilling experiences that promote player well-being. As part of these initiatives, companies such as EA and Nintendo [22], Ubisoft [23] and Riot Games [24] often share data with academics and enable studies into fostering well-being, combating toxicity, and evaluating content moderation. Beyond their main purpose, these studies also expose some of the data that is collected and the information that is detected by these companies lending *transparency* and *auditability* to the industry. These types of cooperations provide a good example of how data transparency can bring value to industry players.

## VI. ADAPTIVE SYSTEMS – OWNERSHIP IN THE AFFECTIVE LOOP

*Adaptation* and affect expression are the steps that close the affective loop. Within games, an affect-based interaction system uses the parameterised output of its affect detection module to adjust the game's parameters to the user experience. The type of adaptation can occur at a macro level through an orchestrator that governs content in the game or the micro level through the behaviour and affect the expression of individual game agents. The goal of the game adaptation might differ depending on the given use case. It might serve to maintain, amplify or change the user's experience, but regardless of the interaction task, the

adaptation module will yield a new set of emotional stimuli, thereby, closing the affective loop (see Fig. 1).

When we look at an adaptive system described by the affective game loop as a whole the question of ownership arises. While AI models can incorporate data from a large number of players, it is unclear how much ownership these players have over these affective models. This issue is even more complex in closed ecosystems that can facilitate co-creation with AI designers. As discussed earlier in Section II, affective computing applications face challenges addressing these questions under the current regulatory frameworks [51], [53], [54]. Even though users should have (at very least) rights to have control over their own data, to portray themselves accurately, to be forgotten [104], and to be self-determined, in reality, information inferred from Big Data is often exempt from the same legal protection afforded to first-hand personal data [31]. To handle the data used to train models and the inferences made by these models on an individual basis Wachter and Mittelstadt propose the concepts of "high-risk inferences" and the "right to reasonable inferences" [31]. The former concept refers to inferences made from Big Data through algorithmic means that are either harmful to the privacy of the user or have low verifiability in what the authors call "important decisions"—with loans and employment brought as examples [31]. The latter concept would enshrine a right that could force data controllers to provide certain information about their inferences:

> "This disclosure would address (1) why certain data form a normatively acceptable basis from which to draw inferences; (2) why these inferences are relevant and normatively acceptable for the chosen processing purpose or type of automated decision; and (3) whether the data and methods used to draw the inferences are accurate and statistically reliable." [31, page 8].

While this proposal—if it were to go into effect—could help provide more robust protection to users, unfortunately, when it comes to the models themselves matters get complicated. At the moment of writing, IP protection virtually takes precedence over the individual's autonomy over their personal data when it comes to the trained models themselves. As the models are considered only "inferred from data" they are further removed from the users whose data is used to create the models [31]. One key challenge with the existing legal framework is that the training data and the model are not as well separated or modular as the guidelines suggest. In many cases, it is possible to reverse engineer the model and extract information about the training set including sensitive personal information about the original subjects [105]. There are algorithmic solutions to address this problem, however. Ongoing research in the field of *machine unlearning* aims to offer methods that attempt to remove knowledge from a trained model as if the given datapoint was never part of the training set [106], [107], [108], [109]. Although early approaches focused on very specific applications—such as *decremental learning* in Support-Vector Machines [110]—recent methods seem to be able to generalise well over different architectures [106], [108], [109]. However, for a practical application of *machine unlearning*, the owner of the model has to retain the raw training data in most cases. Not all methods require this from a technical perspective [109], but

the removal of the datapoint has to be verified to preserve the *predictability* of the system. Although the implementation of unlearning would most certainly pose a computational and organisational overhead, it is still more cost- and resource-efficient than retraining the models from scratch. Even though *machine unlearning* has not been adopted widely yet, contemporary research results show a promising path ahead for mitigating some of the privacy issues in small-scale architectures.

Ethical questions in small-scale models can potentially be addressed through the aforementioned methods, however, far more complex ethical challenges are posed by large-scale world models or *foundation models* [77]. These are large-scale pre-trained models built using hundreds of billions of parameters and massive-scale datasets often scraped from the internet indiscriminately. Foundation models have the potential to provide basic knowledge in a domain or generate content out of the box. While many applications like *AI Dungeon* utilize these foundation models, the lines of *responsibilities* are blurred as a given company has no access to the source code or the original data of these models. As the source of the data is often scraped from the internet, the ethics of constructing such a dataset is also highly questionable. As users are not notified they have no way to revoke their participation or retaliate against their creations used for training these models. In addition, the underlying data is often discarded or kept secret, removing even the moral right to the output of these generators once the model is constructed. Moreover, as these foundation models can only be constructed using an immense amount of resources, large industry players can essentially monopolise the market. This trend can already be seen in language-based models, where the current dominance of the closed-source GPT-3 and—increasingly more popular—GPT-4 [111] models imply that new applications have to subscribe to the black-box rules of that system. Ownership over the input, output and the models themselves is not a trivial problem. The models are owned by their respective companies, and even though users generally retain rights to their input and to the output of the models, there are some major caveats. As an example, the Terms and Conditions of *OpenAI* to their GTP-3 and GPT-4 algorithms[2] which warns users that *OpenAI* themselves retain the rights to both the user input and the system output to improve the system in the future. Nevertheless, users have input confidential information into the system that lead to security leaks. Most recently in the case of *Samsung*, where confidential notes and source code has been leaked through ChatGPT [112]. While users can *opt-out* of this data collection, it is unclear if already submitted data can be removed from the trained model. The ownership issues are further complicated by the secretiveness of the industry stakeholders. Industry players are often invested in creating legal opacity around their systems through restrictive licences and digital rights management tools to restrict the transparency and usage of the software. Although this type of opacity does not stem from the AI architecture itself, it prevents public access to the inner workings of such systems and limits the overview of the AI decision-making process [32].



Fig. 5. In *Dreams* (Sony Interactive Entertainment, 2020), players are able to create their own experiences using a complex editor, reminiscent of the user interface of professional game engines.

Even though the ownership over the models themselves is a central question, we must not forget about the ownership and responsibility over the output of said models either. Who owns the results of a human-AI co-creation process? Looking at contemporary legal frameworks, it is hard to say. Of course, judgement can be passed based on the circumstances and the particularities on a case-by-case basis, but there is no apparent clear line [113] and in most cases, the question is sidestepped entirely by the end-user licence agreement of specific AI-assisted tools. There is no comprehensive framework for either professional creative tools or interactive media meant for entertainment. The matter is further complicated because—even if we focus just on games—it can be hard to make a clear demarcation between a creative tool and curated entertainment. A recent example of this conundrum is Media Molecule's *Dreams* (Sony Interactive Entertainment, 2020), a "game about making games"—see Fig. 5. Although *Dreams* presents itself very much like a game, it perhaps has more in common with game engines, such as *Unity*[3] or *GameMaker*.[4] Nevertheless, until recently games created with *Dreams* were not monetisable by the creators and solely beholden to Sony's PlayStation ecosystem [114]. While Media Molecule maintains that their users retain the rights to their own creations, the options for the users to exercise these rights remain limited.[5] After the runaway success of user-created content such as *Defense of the Ancients*—originally created in the *Warcraft III* (Blizzard Entertainment, 2002) map editor—which lead to a boom of Multiplayer Online Battle Area games it is easy to see why companies are trying to retain as much control over user-generated content as possible. There is a case to be made for the user's moral right over their creations, however. Especially in systems where the player demonstrates considerable creative effort during the co-creation process, they should be able to retain all rights to their own intellectual property. While contemporary examples are still subject to ad-hoc judgement, we expect the right to the output of co-creative systems to become a

---

[2]https://openai.com/policies/terms-of-use

[3]https://unity.com/

[4]https://gamemaker.io/

[5]Media Molecule launched a Beta Evaluation for projects that aim to commercialise their creations off the PlayStation ecosystem. However, this program is not open to everybody and approval is granted on an opaque case-by-case basis. Interestingly—as of writing—games are completely excluded from the program. (Read more at https://docs.indreams.me/en/community/news/dreams-beta-evaluation)

central topic in the near future as AI-powered generative systems become more ubiquitous.

The conversation around co-creation is not just about rights but also responsibilities. Who is responsible for the output of the system when an agent learns to act like a bully, creates offensive and abusive content, or is instructed to generate misinformation? While it is easy—and companies are more than ready—to push the blame on the user, as we demonstrated before in Section III, addressing this issue is not as trivial. The main tools to combat the uncertainty around the models' output preemptively are *transparency* and *predictability*—not just from the AI perspective, but from the larger view of the organisation itself. A delineation between harmful content produced on purpose and as a result of a biased or erroneous algorithm has to rely on clear and transparent guidelines on how people are expected to interact with the system, and what the owners of the system consider malicious use. Predictability of the model output and the larger organisational response to adversarial attacks can facilitate a safer environment for all users involved. The maintenance of the *reliability* and *incorruptibility* of the models should take precedence over the user's input. The *responsibility* of designing and deploying these security measures should, however, fall on the creator of the system. Employing security measures in highly modular software systems is far from trivial given the integration of third-party models, especially out-of-box solutions. Nevertheless, as industrial systems maintain their opaqueness the end users cannot be considered fully autonomous. The extent of their autonomy will always be limited by the design of the application, the complexity of the system, and the limited transparency afforded to them.

The industry has been slow to react to the growing concern about trustworthiness; the landscape is changing, however. An excellent example of recent initiatives is Microsoft's *Xbox Transparency Reports*.[6] In this report, Microsoft publishes explanations and statistics about their content moderation policies to increase both the *reliability* and *transparency* of their ecosystem. Although the first report was just released in 2022, the company pledges to release these transparency reports every 6 months. If successful, a large industry player such as Microsoft can inspire the industry at large to follow suit.

## VII. OTHER ISSUES IN GAME AI ETHICS

There are several issues in the ethics of game AI that do not fit comfortably into our current structure, revolving around the affective loop. Even though these are not the core concern of our paper, and going in-depth on each question would require more space than we have here, we want to at least mention these concerns and provide some pointers for further reading.

All games are, in some way, partial representations of the world and processes therein. One might see games as being "about" certain real-world processes [115]. A key feature of games is also that you learn to perform the actions required to win the game as you play them; well-designed games are typically

pedagogical sequences that introduce gradually harder versions of the same challenge, and it has been argued that this is a key component of why games are fun and appealing [90]. While there are many games whose in-game processes are representations of peaceful real-life activities such as gardening in *FarmVille* (Zynga, 2009) or warehouse stocking in *Sokoban* (Thinking Rabbit, 1982), very many games represent some kind of real-life violence. Games about fighting in various forms are ubiquitous, and have been so since the birth of the medium; many, perhaps most, video games have "hit" or "shoot" as one of their most important mechanics. This might be because it is comparatively easy to design engaging games around fighting compared to any other purpose; regardless, combat is pervasive in games since the era of *Pac-Man* (Namco, 1980) and its ghost-eating capability.

One potential issue of this is whether playing video games inspires, encourages, or teaches violent behaviour. There has been a debate around the effects of video games on violent behaviour for at least three decades, and many studies of varying quality. To some extent, this debate and field of inquiry can be subsumed under the broader question of media effects, where the effects of other media (such as TV) on violence and other behavioural aspects have been studied for a longer time. Although the idea that video games in some way lead to violent behaviour has some plausibility as video games teach some kind of skills, thorough and well-performed studies have largely failed to find any causal link [116]. Even if such an unknown link might be there, AI models can now assist in the isolation of such violence of toxic behaviours [21].

Given the very high computing demands of much current AI research and the advantages of having access to large datasets, it is worth pondering which modern game AI methods will benefit most. It is possible that modern game AI will exacerbate the divide between large developers with deep pockets, multiple titles and existing user bases and small independent game developers. If models trained on user data stay proprietary, the large developers will have a considerable additional advantage over small creators.

Concerns about fairness and bias are ubiquitous in machine learning, and as discussed previously, these concerns are very real for AI in games as well. It is often claimed that biased models come out of biased teams, in other words, that the composition of the human workforce defining and developing the AI solution impacts bias. This is certainly a concern in the game industry, which appears to be at least as demographically imbalanced as the rest of the tech industry [117], [118].

Finally, a far-reaching potential ethical concern is that we one day develop artificial general intelligence, that is as capable (as we are) across a large range of areas and tasks. Games could have played a critical role in that development. Such entities might become very influential in human affairs, and may also gain the ability to improve themselves, potentially leading to what has been termed an *intelligence explosion* [119]. If that happens, the alignment problem becomes acute: making sure that the goals and principles of such an entity are aligned with human society. Given that video games are abundantly used in AI research, it is worth pondering what impact training on video games might have on the ethics of a potential superintelligence.

---

[6]https://www.xbox.com/en-GB/legal/xbox-transparency-report

## VIII. CONCLUSION

This survey paper discussed thoroughly the various ethical aspects of artificial intelligence *in* and *for* games. We opted to view the most critical of these aspects under the affective game loop concept [9]. Based on that concept we reviewed the current game AI state-of-the-art and the game industry state-of-practice with respect to player experience *elicitation*, *sensing*, *detection* and finally *adaptation*. We raised a number of ethical dimensions and concerns and the current (lack of) measures and tools available to address them. We also made a number of recommendations and suggested future steps for making ethics an integral part of AI and games research and innovation. The dialogue between the game industry and academic stakeholders is currently active across the various conferences (e.g., GDC, IEEE CoG, FDG, CHI Play), seminars and summer schools (e.g., AI and Games Summer School) with a focus on the area. Moreover, the ethical aspects of AI in games and media at large are currently a top priority item in the agenda of policymakers (e.g., the European Commission) manifested through research and innovation projects[7,8] and policies such as the AI act.[9] We expect affective computing researchers to take a leading role in these efforts. Affective computing is uniquely positioned as a multidisciplinary field between sensor technology, AI, and applied psychology; hence it offers a comprehensive overview of most of the issues this survey has touched upon. Although in many cases the response to emerging issues has to be regulatory instead of technical, affective computing can still provide a shared language between the fields involved and help highlighting potential issues. This paper aims to further facilitate and moderate this dialogue among all stakeholders involved—AI and affective computing researchers and practitioners, game developers, and ultimately players—with the hope that ethical awareness is increased and that necessary action is taken for the mutual benefit of players and their games.

## REFERENCES

[1] J. P. Zagal, S. Björk, and C. Lewis, "Dark patterns in the design of games," in *Proc. 8th Int. Conf. Foundations Digit. Games*, SASDG, 2013, pp. 39–46.

[2] D. L. King, P. H. Delfabbro, S. M. Gainsbury, M. Dreier, N. Greer, and J. Billieux, "Unfair play? Video games as exploitative monetized services: An examination of game patents from a consumer protection perspective," *Comput. Hum. Behav.*, vol. 101, pp. 131–143, 2019.

[3] B. Mikkelsen, C. Holmgard, and J. Togelius, "Ethical considerations for player modeling," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 975–982.

[4] M. Cook, "Ethical procedural generation," in *Procedural Generation in Game Design*, Natick, MA, USA/Boca Raton, FL, USA: AK Peters/CRC Press, 2017, pp. 43–54.

[5] M. Seif El-Nasr and E. Kleinman, "Data-driven game development: Ethical considerations," in *Proc. Int. Conf. Found. Digit. Games*, 2020, pp. 1–10.

[6] M. Cook, "The social responsibility of game AI," in *Proc. IEEE Conf. Games*, 2021, pp. 1–8.

[7] G. N. Yannakakis, P. Spronck, D. Loiacono, and E. André, "Player modeling," *Artif. Comput. Intell. Games*, ser. Dagstuhl Follow-Ups. Schloss Dagstuhl — Leibniz-Zentrum für Informatik, vol. 6, 2013, pp. 45–59.

[8] C. Holmgård, A. Liapis, J. Togelius, and G. N. Yannakakis, "Evolving personas for player decision modeling," in *Proc. IEEE Conf. Comput. Intell. Games*, 2014, pp. 1–8.

[9] G. N. Yannakakis and A. Paiva, "Emotion in games," *Handbook Affect. Comput.*, vol. 2014, pp. 459–471, 2014.

[10] G. N. Yannakakis and J. Togelius, "Experience-driven procedural content generation," *IEEE Trans. Affect. Comput.*, vol. 2, no. 3, pp. 147–161, Third Quarter 2011.

[11] A. Liapis, G. N. Yannakakis, M. J. Nelson, M. Preuss, and R. Bidarra, "Orchestrating game generation," *IEEE Trans. Games*, vol. 11, no. 1, pp. 48–68, Mar. 2019.

[12] J. Togelius and G. N. Yannakakis, "General general game AI," in *Proc. IEEE Conf. Comput. Intell. Games*, 2016, pp. 1–8.

[13] G. N. Yannakakis and J. Hallam, "Real-time game adaptation for optimizing player satisfaction," *IEEE Trans. Comput. Intell. AI Games*, vol. 1, no. 2, pp. 121–133, Jun. 2009.

[14] N. Shaker, G. Yannakakis, and J. Togelius, "Towards automatic personalized content generation for platform games," in *Proc. AAAI Conf. Artif. Intell. Interactive Digit. Entertainment*, 2010, pp. 63–68.

[15] M. Van Rooij, A. Lobel, O. Harris, N. Smit, and I. Granic, "DEEP: A biofeedback virtual reality game for children at-risk for anxiety," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, 2016, pp. 1989–1997.

[16] D. Villani, C. Carissoli, S. Triberti, A. Marchetti, G. Gilli, and G. Riva, "Videogames for emotion regulation: A systematic review," *Games Health J.*, vol. 7, no. 2, pp. 85–99, 2018.

[17] D. Melhart, A. Liapis, and G. N. Yannakakis, "Towards general models of player experience: A study within genres," in *Proc. IEEE Conf. Games*, 2021, pp. 01–08.

[18] N. Bostrom and E. Yudkowsky, "The ethics of artificial intelligence," *Cambridge Handbook Artif. Intell.*, vol. 1, pp. 316–334, 2014.

[19] I. Subhan, "18 European countries call for better regulation of loot boxes following new report," 2022. [Online]. Available: https://www.eurogamer.net/18-european-countries-call-for-better-regulation-of-loot-boxes-following-new-report

[20] A. Hon, *You've Been Played: How Corporations, Governments and Schools Use Games to Control Us All*. London, U.K.: Swift Press, 2022.

[21] A. Canossa, D. Salimov, A. Azadvar, C. Harteveld, and G. Yannakakis, "For honor, for toxicity: Detecting toxic behavior through gameplay," *Proc. ACM Hum.-Comput. Interaction*, vol. 5, pp. 1–29, 2021.

[22] N. Johannes, M. Vuorre, and A. K. Przybylski, "Video game play is positively correlated with well-being," *Roy. Soc. Open Sci.*, vol. 8, no. 2, 2021, Art. no. 202049.

[23] D. Melhart, A. Azadvar, A. Canossa, A. Liapis, and G. N. Yannakakis, "Your gameplay says it all: Modelling motivation in tom clancy's the division," in *Proc. IEEE Conf. Games*, 2019, pp. 1–8.

[24] C. Monge and T. O'Brien, "Effects of individual toxic behavior on team performance in league of legends," *Media Psychol.*, vol. 25, no. 1, pp. 82–105, 2022.

[25] J. J. Bryson and P. P. Kime, "Just an artifact: Why machines are perceived as moral agents," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1641–1646.

[26] H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser, and Q. Yang, "Building ethics into artificial intelligence," 2018, *arXiv:1812.02953*.

[27] J. Crowley et al., "Toward AI systems that augment and empower humans by understanding us, our society and the world around us," *Rep. 761758 EU Project HumaneAI*, pp. 1–32, 2019.

[28] S. Larsson and F. Heintz, "Transparency in artificial intelligence," *Internet Policy Rev.*, vol. 9, no. 2, pp. 1–16, 2020.

[29] S. Thiebes, S. Lins, and A. Sunyaev, "Trustworthy artificial intelligence," *Electron. Markets*, vol. 31, no. 2, pp. 447–464, 2021.

[30] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (XAI): A survey," 2020, *arXiv:2006.11371*.

[31] S. Wachter and B. Mittelstadt, "A right to reasonable inferences: Rethinking data protection law in the age of big data and AI," *Columnia Bus. Law Rev.*, no. 2, pp. 494–620, 2019.

[32] L. Beckman, J. Hultin Rosenberg, and K. Jebari, "Artificial intelligence and democratic legitimacy. the problem of publicity in public authority," *AI Soc.*, vol. 1, pp. 1–10, 2022.

---

[7] https://www.ai4media.eu/

[8] https://learnml.eu/

[9] https://artificialintelligenceact.eu/

[33] H. Liu et al., "Trustworthy AI: A computational perspective," 2021, *arXiv:2107.06641*.

[34] T. Gebru, "Race and gender," in *The Oxford Handbook of Ethics of AI*, Oxford, U.K.: Oxford Univ. Press, 2020, pp. 251–269.

[35] V. Vakkuri, K.-K. Kemell, J. Kultanen, M. Siponen, and P. Abrahamsson, "Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study," 2019, *arXiv:1906.07946*.

[36] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems," 2019. [Online]. Available: https://standards.ieee.org/industry-connections/ec/autonomous-systems

[37] L. Floridi, "Establishing the rules for building trustworthy AI," *Nature Mach. Intell.*, vol. 1, no. 6, pp. 261–262, 2019.

[38] N. A. Smuha, "The EU approach to ethics guidelines for trustworthy artificial intelligence," *Comput. Law Rev. Int.*, vol. 20, no. 4, pp. 97–106, 2019.

[39] H. Yu, C. Miao, C. Leung, and T. J. White, "Towards AI-powered personalization in MOOC learning," *npj Sci. Learn.*, vol. 2, no. 1, 2017, Art. no. 15.

[40] V. Dignum, "Ethics in artificial intelligence: Introduction to the special issue," *Ethics Inf. Technol.*, vol. 20, pp. 1–3, 2018.

[41] M. Rovatsos, "We may not cooperate with friendly machines," *Nature Mach. Intell.*, vol. 1, no. 11, pp. 497–498, 2019.

[42] F. Ishowo-Oloko, J.-F. Bonnefon, Z. Soroye, J. Crandall, I. Rahwan, and T. Rahwan, "Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation," *Nature Mach. Intell.*, vol. 1, no. 11, pp. 517–521, 2019.

[43] S. Russell, S. Hauert, R. Altman, and M. Veloso, "Ethics of artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 415–416, 2015.

[44] S. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*. London, U.K.: Penguin, 2019.

[45] T. Hagendorff, "The ethics of ai ethics: An evaluation of guidelines," *Minds Machines*, vol. 30, no. 1, pp. 99–120, 2020.

[46] J. Zhu and M. S. El-Nasr, "Open player modeling: Empowering players through data transparency," 2021, *arXiv:2110.05810*.

[47] M. Sicart, *The Ethics of Computer Games*. Cambridge, MA, USA: MIT Press, 2011.

[48] J. Zhu, A. Liapis, S. Risi, R. Bidarra, and G. M. Youngblood, "Explainable ai for designers: A human-centered perspective on mixed-initiative co-creation," in *Proc. IEEE Conf. Comput. Intell. Games*, 2018, pp. 1–8.

[49] A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020.

[50] D. Hooshyar, E. Bardone, N. E. Mawas, and Y. Yang, "Transparent player model: Adaptive visualization of learner model in educational games," in *Proc. Int. Conf. Innov. Technol. Learn.*, Springer, 2020, pp. 349–357.

[51] J. L. Kröger, P. Raschke, J. P. Campbell, and S. Ullrich, "Surveilling the gamers: Privacy impacts of the video game industry," *Entertainment Comput.*, vol. 44, 2023, Art. no. 100537.

[52] P. Voigt and A. Von dem Bussche, "The EU general data protection regulation (GDPR)," in *A Practical Guide*, 1st Ed., Cham, Switzerland: Springer International Publishing, 2017, Art. no. 3152676.

[53] P. Jost and M. Lampert, "Two years after: A scoping review of GDPR effects on serious games research ethics reporting," in *Proc. Int. Conf. Games Learn. Alliance*, Springer, 2020, pp. 372–385.

[54] A. Häuselmann, "Fit for purpose? Affective computing meets EU data protection law," *Int. Data Privacy Law*, vol. 11, pp. 245–256, 2021.

[55] S. McCrea, G. Geršak, and D. Novak, "Absolute and relative user perception of classification accuracy in an affective video game," *Interacting Comput.*, vol. 29, no. 2, pp. 271–286, 2017.

[56] D. Zendle and P. Cairns, "Video game loot boxes are linked to problem gambling: Results of a large-scale survey," *PLoS One*, vol. 13, no. 11, 2018, Art. no. e0206767.

[57] A. Drummond, J. D. Sauer, L. C. Hall, D. Zendle, and M. R. Loudon, "Why loot boxes could be regulated as gambling," *Nature Hum. Behav.*, vol. 4, no. 10, pp. 986–988, 2020.

[58] E. Gibson, M. Griffiths, F. Calado, and A. Harris, "The relationship between videogame micro-transactions and problem gaming and gambling: A systematic review," *Comput. Hum. Behav.*, vol. 131, 2022, Art. no. 107219.

[59] S. E. Hodge, M. Vykoukal, J. McAlaney, R. D. Bush-Evans, R. Wang, and R. Ali, "What's in the box? exploring UK players' experiences of loot boxes in games; The conceptualisation and parallels with gambling," *PLoS One*, vol. 17, no. 2, 2022, Art. no. e0263567.

[60] E. Petrovskaya and D. Zendle, "The battle pass: A mixed-methods investigation into a growing type of video game monetisation," *OSF Preprints*, 2020.

[61] D. Joseph, "Battle pass capitalism," *J. Consum. Culture*, vol. 21, no. 1, pp. 68–83, 2021.

[62] D. L. King, A. M. Russell, P. H. Delfabbro, and D. Polisena, "Fortnite microtransaction spending was associated with peers' purchasing behaviors but not gaming disorder symptoms," *Addictive Behaviors*, vol. 104, 2020, Art. no. 106311.

[63] Á. Periáñez, A. Saas, A. Guitart, and C. Magne, "Churn prediction in mobile social games: Towards a complete assessment using survival ensembles," in *Proc. Int. Conf. Data Sci. Adv. Analytics*, 2016, pp. 564–573.

[64] P. P. Chen, A. Guitart, A. F. del Río, and A. Periáñez, "Customer lifetime value in video games using deep learning and parametric models," in *Proc. IEEE Int. Conf. Big Data*, 2018, pp. 2134–2140.

[65] M. Griffiths, "A 'components' model of addiction within a biopsychosocial framework," *J. Substance Use*, vol. 10, no. 4, pp. 191–197, 2005.

[66] M. Xi, Z. Luo, N. Wang, and J. Yin, "A latent feelings-aware RNN model for user churn prediction with behavioral data," 2019, *arXiv:1911.02224*.

[67] K. K. Mak, K. Lee, and C. Park, "Applications of machine learning in addiction studies: A systematic review," *Psychiatry Res.*, vol. 275, pp. 53–60, 2019.

[68] R. Valentine, "Ea faces yet another class-action lawsuit connected to loot boxes," 2020. [Online]. Available: https://www.gamesindustry.biz/ea-faces-yet-another-class-action-lawsuit-over-alleged-use-of-dynamic-difficulty-adjustment

[69] J. Batchelor, "'dynamic difficulty' loot box lawsuit against ea dropped," 2021. [Online]. Available: https://www.gamesindustry.biz/dynamic-difficulty-loot-box-lawsuit-against-ea-dropped

[70] F. Tencé, C. Buche, P. De Loor, and O. Marc, "The challenge of believability in video games: Definitions, agents models and imitation learning," 2010, *arXiv:1009.0451*.

[71] K. Jorgensen and F. Karlsen, *Transgression in Games and Play*. Cambridge, MA, USA: MIT Press, 2019.

[72] E. Sheng, K.-W. Chang, P. Natarajan, and N. Peng, "The woman worked as a babysitter: On biases in language generation," 2019, *arXiv:1909.01326*.

[73] L. Lucy and D. Bamman, "Gender and representation bias in GPT-3 generated stories," in *Proc. 3rd Workshop Narrative Understanding*, 2021, pp. 48–55.

[74] P. Budzianowski and I. Vulić, "Hello, it's GPT-2–how can i help you? Towards the use of pretrained language models for task-oriented dialogue systems," 2019, *arXiv:1907.05774*.

[75] L. Floridi and M. Chiriatti, "GPT-3: Its nature, scope, limits, and consequences," *Minds Machines*, vol. 30, no. 4, pp. 681–694, 2020.

[76] T. Simonite, "It began as an AI-fueled dungeon game. it got much darker," 2021. [Online]. Available: https://www.wired.com/story/ai-fueled-dungeon-game-got-much-darker/

[77] R. Bommasani et al., "On the opportunities and risks of foundation models," 2021, *arXiv:2108.07258*.

[78] C. Reynolds and R. Picard, "Affective sensors, privacy, and ethical contracts," in *Proc. Extended Abstr. Hum. Factors Comput. Syst.*, 2004, pp. 1103–1106.

[79] R. Cowie, "Ethical issues in affective computing," in *The Oxford handbook of affective computing*, Oxford, U.K.: Oxford Univ. Press, 2015, pp. 334–348.

[80] J. L. Kröger, P. Raschke, J. P. Campbell, and S. Ullrich, "Surveilling the gamers: Privacy impacts of the video game industry," *Entertainment Comput.*, vol. 44, pp. 1–23, 2021.

[81] S. Makarovych, A. Canossa, J. Togelius, and A. Drachen, "Like a dna string: Sequence-based player profiling in tom clancy's the division," in *Proc. Artif. Intell. Interactive Digit. Entertainment Conf.*, York, 2018, pp. 152–158.

[82] C. M. Myers, L. F. Laris Pardo, A. Acosta-Ruiz, A. Canossa, and J. Zhu, ""try, try, try again:" Sequence analysis of user interaction data with a voice user interface," in *Proc. 3rd Conf. Conversational User Interfaces*, 2021, pp. 1–8.

[83] L. M. Vizer, L. Zhou, and A. Sears, "Automated stress detection using keystroke and linguistic features: An exploratory study," *Int. J. Hum.-Comput. Stud.*, vol. 67, no. 10, pp. 870–886, 2009.

[84] J. L. Kröger, O. H.-M. Lutz, and P. Raschke, "Privacy implications of voice and speech analysis–information disclosure by inference," in *Proc. IFIP Int. Summer Sch. Privacy Identity Manage.*, Springer, 2019, pp. 242–258.

[85] M. S. El-Nasr, A. Drachen, and A. Canossa, *Game Analytics*. Berlin, Germany: Springer, 2016.

[86] S. C. Bakkes, P. H. Spronck, and G. van Lankveld, "Player behavioural modelling for video games," *Entertainment Comput.*, vol. 3, no. 3, pp. 71–79, 2012.

[87] R. Hare and Y. Tang, "Player modelling and adaptation methods within adaptive serious games," in *Proc. Int. Conf. Cyber- Phys. Social Intell.*, 2021, pp. 1–6.

[88] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.

[89] H. Wang and C.-T. Sun, "Game reward systems: Gaming experiences and social meanings," in *Proc. DiGRA Int. Conf.: Think Des. Play*, DiGRA/Utrecht School of the Arts, 2011, pp. 1–15.

[90] R. Koster, *Theory of Fun for Game Design*. Sebastopol, CA, USA: O'Reilly Media, Inc., 2013.

[91] S. Abuhamdeh, M. Csikszentmihalyi, and B. Jalal, "Enjoying the possibility of defeat: Outcome uncertainty, suspense, and intrinsic motivation," *Motivation Emotion*, vol. 39, no. 1, pp. 1–10, 2015.

[92] B. Cowley, D. Charles, M. Black, and R. Hickey, "Toward an understanding of flow in video games," *Comput. Entertainment*, vol. 6, no. 2, pp. 1–27, 2008.

[93] S. Rigby and R. M. Ryan, *Glued to Games: How Video Games Draw Us in and Hold Us Spellbound: How Video Games Draw Us in and Hold Us Spellbound*. Santa Barbara, CA, USA: AbC-CLIo, 2011.

[94] J. Huizinga, *Homo Ludens: A Study of the Play Element in Culture*. Boston, MA, USA: Beacon Press, 1955.

[95] K. S. Tekinbas and E. Zimmerman, *Rules of Play: Game Design Fundamentals*. Cambridge, MA, USA: MIT Press, 2003.

[96] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Portland, OR, USA: Broadway Books, 2016.

[97] K. Lum and W. Isaac, "To predict and serve?," *Significance*, vol. 13, no. 5, pp. 14–19, 2016.

[98] A. Yapo and J. Weiss, "Ethical implications of bias in machine learning," in *Proc. 51st Hawaii Int. Conf. Syst. Sci.*, 2018, pp. 1–8.

[99] T. Gebru et al., "Datasheets for datasets," *Commun. ACM*, vol. 64, no. 12, pp. 86–92, 2021.

[100] D. Roselli, J. Matthews, and N. Talagala, "Managing bias in AI," in *Proc. World Wide Web Conf.*, 2019, pp. 539–544.

[101] J. Vincent, "Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech," 2018. [Online]. Available: https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai

[102] M. Fahey, "Autistic boy branded a cheater by xbox live [update]," 2013. [Online]. Available: https://kotaku.com/autistic-boy-branded-a-cheater-by-xbox-live-update-5743970

[103] B. Ashcraft, "Korean woman kicks ass at overwatch, gets accused of cheating [update]," 2016. [Online]. Available: https://kotaku.com/korean-woman-kicks-ass-at-overwatch-gets-accused-of-ch-1782343447

[104] E. F. Villaronga, P. Kieseberg, and T. Li, "Humans forget, machines remember: Artificial intelligence and the right to be forgotten," *Comput. Law Secur. Rev.*, vol. 34, no. 2, pp. 304–313, 2018.

[105] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction APIs," in *Proc. 25th USENIX Secur. Symp.*, 2016, pp. 601–618.

[106] C. Guo, T. Goldstein, A. Hannun, and L. Van Der Maaten, "Certified data removal from machine learning models," 2019, *arXiv:1911.03030*.

[107] L. Graves, V. Nagisetty, and V. Ganesh, "Amnesiac machine learning," 2020, *arXiv:2010.10981*.

[108] L. Bourtoule et al., "Machine unlearning," in *Proc. IEEE Symp. Secur. Privacy*, 2021, pp. 141–159.

[109] A. Sekhari, J. Acharya, G. Kamath, and A. T. Suresh, "Remember what you want to forget: Algorithms for machine unlearning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 18075–18086.

[110] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 388–394.

[111] OpenAI, "GPT-4 technical report," 2023, *arXiv:2303.08774*.

[112] L. Maddison, "Samsung workers made a major error by using ChatGPT," 2023. [Online]. Available: https://www.techradar.com/news/samsung-workers-leaked-company-secrets-by-using-chatgpt

[113] J. K. Eshraghian, "Human ownership of artificial creativity," *Nature Mach. Intell.*, vol. 2, no. 3, pp. 157–160, 2020.

[114] J. Castello, "PS4 game dreams is an amazing creation tool with an exposure problem," 2020. [Online]. Available: https://www.theverge.com/2020/2/14/21136244/dreams-ps4-game-creation-tool-exposure-problem-curation-media-molecule

[115] I. Bogost, *Persuasive Games: The Expressive Power of Videogames*. Cambridge, MA, USA: MIT Press, 2010.

[116] J. Breuer, J. Vogelgesang, T. Quandt, and R. Festl, "Violent video games and physical aggression: Evidence for a selection effect among adolescents," *Psychol. Popular Media Culture*, vol. 4, no. 4, pp. 305–328, 2015.

[117] E. N. Bailey, K. Miyata, and T. Yoshida, "Gender composition of teams and studios in video game development," *Games Culture*, vol. 16, no. 1, pp. 42–64, 2021.

[118] C. J. Passmore, R. Yates, M. V. Birk, and R. L. Mandryk, "Racial diversity in indie games: Patterns, challenges, and opportunities," in *Proc. Extended Abstr. Publication Annu. Symp. Comput.-Hum. Interaction Play*, 2017, pp. 137–151.

[119] N. Bostrom, "How long before superintelligence?," *Int. J. Futures Stud.*, vol. 2, pp. 12–17, 1998.

**David Melhart** received the MA degree in cognition and communication from the University of Copenhagen, in 2016, and the PhD degree in game research from the University of Malta, in 2021. He is a senior member of technical staff with modl.ai, and a postdoctoral researcher with the Institute of Digital Games, University of Malta. His research focuses on machine learning, affective computing, and games user modelling. He has been the Communication chair of FDG 2020, a recurring organiser and Publicity chair of the *Summer School series on Artificial Intelligence and Games* (2018–2023), the Workshop and Panels chair of FDG 2023, editorial assistant of *IEEE Transactions on Games*, guest associate editor of *User States in Extended Reality Media Experiences for Entertainment Games* Special Issue of *Frontiers in Virtual Reality and Human Behaviour*, and review editor of *Frontiers in Human-Media Interaction*.

**Julian Togelius** received the BA degree from Lund University, the MSc degree from the University of Sussex, and the PhD degree from the University of Essex. He is a co-founder and research director of modl.ai, and an associate professor with the Department of Computer Science and Engineering, New York University. He works on artificial intelligence for games and on games for artificial intelligence. His current main research directions involve procedural content generation in games, general video game playing, player modelling, and fair and relevant benchmarking of AI through game-based competitions. Additionally, he works on topics in evolutionary computation, quality-diversity algorithms, and reinforcement learning. From 2018 to 2021, he was an editor-in-chief of *IEEE Transactions on Games*. He has previously worked with IDSIA in Lugano and with the IT University of Copenhagen.

**Benedikte Mikkelsen** received the BA degree in architecture from The Royal Danish Academy of Fine Arts, and the MSc degree in media technology from the IT University of Copenhagen. She is a co-founder and chief product officer of modl.ai. She formed her first game development company Duck and Cover Games, focusing on learning game development, with Christoffer Holmgård, in 2008, where they shipped multiple bespoke game titles for large public and private actors in Denmark. Subsequently, she worked with user experience and web application development, among other things developing design-driven integrated web and mobile platforms for actors in the public sector. She has experience working in research and development from international European Union Research projects focused on building non-expert end user behavioural modelling tools leveraging game telemetry data.

**Christoffer Holmgård** received the BA degree in psychology from the University of Copenhagen and the MSc degree in media technology from the IT University of Copenhagen, and the PhD degree in artificial intelligence and procedural content generation from the IT University of Copenhagen, in 2015, and further earned a post-doctorate in Game Engineering from New York University. He is a co-founder and chief executive officer of modl.ai. He worked in statistics, psychometrics, and organisational psychology with the Royal Danish Defence College from 2004 until 2011, maintaining the Danish national draft board intelligence test, conducting psychometric assessment of special forces operatives, fighter pilots, and officers, and supporting Danish veterans. Before starting modl.ai, he served as a tenure-track assistant professor with Northeastern University and the head of their master's program in Game Science and Design. He formed his first game development company, in 2008, Duck and Cover Games, focusing on learning games, together with Benedikte Mikkelsen. In 2011, he co-founded the award winning game studio Die Gute Fabrik, leading the studio as managing director, today serving as the chair of the board. During his tenure, the studio shipped multiple games on PC, Mac, mobile, and console platforms, securing more than 22 industry nominations and awards, including the GDC Innovation Award and the IndieCade Grand Jury Award.

**Georgios N. Yannakakis** (Senior Member, IEEE) received the PhD degree in informatics from the University of Edinburgh, in 2006. He is a co-founder of modl.ai, and a professor and director of the Institute of Digital Games, University of Malta (UM). Prior to joining UM, in 2012 he was an associate professor with the Center for Computer Games Research, IT University of Copenhagen. He does research with the crossroads of artificial intelligence, affective computing, games and computational creativity. He has published more than 350 papers in the aforementioned fields and his work has been cited broadly. His research has been supported by numerous national and European grants (including a Marie Skłodowska-Curie Fellowship). He is currently an editor in chief of *IEEE Transactions on Games*, an associate editor of *IEEE Transactions on Evolutionary Computation*, and used to be an associate editor of *IEEE Transactions on Affective Computing* and *IEEE Transactions on Computational Intelligence and AI in Games* journals. Among the several rewards he has received for his papers he is a recipient of *IEEE Transactions on Affective Computing Most Influential Paper Award* and *IEEE Transactions on Games Outstanding Paper Award*.