AI and Ethics (2023) 3:407–417
https://doi.org/10.1007/s43681-022-00158-4

ORIGINAL RESEARCH

Agency in augmented reality: exploring the ethics of Facebook's AI-powered predictive recommendation system

Andreas Schönau[1,2]

Received: 21 February 2022 / Accepted: 4 April 2022 / Published online: 25 April 2022

Abstract
The development of predictive algorithms for personalized recommendations that prioritize ads, filter content, and tailor our decision-making processes will increasingly impact our society in the upcoming years. One example of what this future might hold was recently presented by Facebook Reality Labs (FRL) who work on augmented reality (AR) glasses powered by contextually aware AI that allows the user to "communicate, navigate, learn, share, and take action in the world" (Facebook Reality Labs 2021). A major feature of those glasses is "the intelligent click" that presents action prompts to the user based on their personal history and previous choices. The user can accept or decline those suggested action prompts depending on individual preferences. Facebook/Meta presents this technology as a gateway to "increased agency". However, Facebook's claim presumes a simplistic view of agency according to which our agentive capacities increase parallel to the ease in which our actions are carried out. Technologies that structure people's lives need to be based on a deeper understanding of agency that serves as the conceptual basis in which predictive algorithms are developed. With the goal of mapping this emerging terrain, the aim of this paper is to offer a thorough analysis of the agency-limiting risks and the agency-enhancing potentials of Facebook's "intelligent click" feature. Based on a concept of agency by Dignum (Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer International Publishing, Cham, 2019), the three agential dimensions of autonomy (acting independently), adaptability (reacting to changes in the environment), and interactivity (interacting with other agents) are analyzed towards our ability to make self-determining choices.

Keywords AI ethics · Augmented reality · Predictive recommendation algorithms · Facebook · Meta · Agency

1 Introduction

In the not so distant future, you might have technologies integrated into your everyday routine that are powered by Artificial Intelligence (AI) in ways that not only impact but fundamentally shape your life. As one of those future technologies, imagine a portable device that helps you to structure your daily routine by analyzing your behavior through microphones and cameras that are capturing everything around you. Based on predictive algorithms that distinguish between things you probably like and things you probably don't like, you get recommendations about what to do next (e.g., "Do you want to go for a walk?") or prompts about tasks the system could take over for you (e.g., "Do you want me to run the dishwasher for you?"). Imagine that those recommendations about activities are not restricted to your home environment but that they encompass everything you experience from the moment you turn the device on—whether you go outside for your weekly run, prepare a presentation you are eager to finish for work, or discuss with a friend about where to meet.

While this is a hypothetical future that has not yet become reality, ongoing research conducted by Facebook Reality Labs (FRL) aims at developing a predictive AI system that is similar to the scenario depicted above. Facebook is trying to reach that goal with augmented reality (AR) glasses powered by contextually aware AI that allows the user to "communicate, navigate, learn, share, and take action in the world" [10]. One of the major features of these glasses is "the intelligent click" that presents action prompts to the user based on their personal history and previous choices. The user can

Andreas Schönau
schoenau@uw.edu

[1] Department of Philosophy, University of Washington, Seattle, WA, USA

[2] Center for Neurotechnology, University of Washington, Seattle, WA, USA

accept or decline those suggested action prompts depending on their individual preference. Facebook presents this technology as a facilitator for a new generation of communication, access, and navigation that "leads to this phenomenon of increased agency of you feeling like a level of control you've never had before."[1]

While a future in which AI assists our actions can be somewhat enticing, it is also daunting insofar that it is possible that algorithms could take over our agentive capacities instead of empowering them, leaving an array of ethical uncertainties: In what ways are our autonomous choices constrained if a system structures our weekly routine? Can we become over-reliant on AI for problem solving? How genuine are our interactions with friends if they are based on advice given by an algorithm?

These questions raise significant concerns given that Facebook's approach leaves it conceptually unclear in what way our agency is "increased" if the device is prompting us to perform an action. As of now, Facebook seems to presume a simplistic view of agency, wherein our agentive capacities increase parallel to the ease in which our actions are carried out (a thorough explanation of FRL's intent will be offered at the end of Sect. 2). However, relying on such an underdeveloped understanding of agency as the conceptual basis on which predictive algorithms are introduced in future technologies risks removing human responsibility for choices, limiting people's decision-making capacities, and devaluing human skills [8, 11]. To avoid these and other potential pitfalls, technologies and algorithms that structure people's lives need to be based on an understanding of agency that takes those complex interrelations into account.

With the goal of mapping this emerging terrain, the aim of this paper is to offer an analysis of the agency-limiting risks and the agency-enhancing potentials of predictive recommendation systems using Facebook's "intelligent click" algorithm as an illustrative example. This analysis can be used as a conceptual basis for ensuring that the integration of predictive algorithms into people's everyday lives are developed in an ethically responsible way.

The paper is structured as follows: Section 2 starts with an overview of the ways in which AI already structures our lives using Netflix's content management system as an illustrative example demonstrating how user preferences can be guided by algorithms. Being confronted with the already existing algorithms that impact our daily lives reveals the necessity to start an analysis about predictive recommendation algorithms before more sophisticated applications hit the market that cover more private areas of our lives. As an example of what this future technology could look like, augmented reality (AR) is described in more detail, as well as FRL's plan of using the "intelligent click" as their major feature that is facilitated by predictive algorithms. Section 3 introduces a theoretically useful conception of agency based on recent work by Dignum [8]. According to Dignum, agency in the context of AI consists of three dimensions: autonomy (i.e., acting independently), adaptability (i.e., reacting to changes in the environment), and interactivity (i.e., interacting with other agents). These three dimensions will be defined and discussed in their ethical weight for predictive recommendation systems. Section 4 offers a contextual analysis by looking at the ways in which the "intelligent click" feature of the FRL–AR glasses could extend or limit a user's agency. Since this technology is still in development and, naturally, lacks any empirical case studies, three hypothetical case scenarios are introduced to explore how each of the three identified aspects of agency—autonomy, adaptability, and interactivity—could be at risk of being compromised by the way in which the predictive recommendation algorithm functions. A conceptual analysis follows the introduction of each of those scenarios to identify potential issues related to compromised agency but also to show how the system could be developed in ways that aid or expand a user's agency.

2 From present to future: the need for ethical oversight of predictive algorithms in current and future technologies

Algorithms that give us recommendations based on our behavior are influencing our choices, contributing to the formation of new habits, and structuring the way we spend our time—all while they keep us engaged with the system for which they are designed. Thinking about the future of AI and the ways it is going to impact our individual lives makes it easy to forget that there are already numerous technologies that use algorithms to recommend content for us to consume. Among many others, a good example is the way Netflix uses personalized predictive algorithms to structure content presented to their users. Netflix is an online streaming service that offers instant access to movies and television shows with

[1] In a recent interview with the verge [24], Facebook's CEO Mark Zuckerberg revealed more information about the companies' long term goals by explaining his vision of generating what he calls a "metaverse", i.e. a space in which the physical and the virtual world come together to build their own economy. Zuckerberg describes this as an "embodied internet where instead of just viewing content—you are in it." This vision became more concrete with the recent announcement of the company's name change from Facebook to Meta at their annual meeting "Connect 2021" [22].

139 million subscribers in over 190 countries as of 2019, adding up to 140 million hours of content every day [7].[2]

As a video-on-demand service, Netflix allows their users to choose among a variety of videos to watch. To keep their audience engaged with the platform, Netflix uses 7 different types of recommendation systems powered by different algorithms that, together, shape every aspect of the user experience, ranging from the genre identified as most engaging to the individual user (e.g., "suspenseful movies" vs. "trending now" vs. "because you watched"), the order of the videos in the "continue watching" row, to the information displayed about videos on screen (e.g., thumbnails, summary of the show): "For example, evidence algorithms decide whether to show that a certain movie won an Oscar or instead show the member that the movie is similar to another video recently watched by that member; they also decide which image out of several versions use to best support a given recommendation" [13]. Taken together, predictive recommendations are a crucial element for the company's success, since they are responsible for 80% of the hours streamed on Netflix [13].

For streaming services such as Netflix, personalized algorithms are used to provide the customer with what they are most likely to enjoy, which in turn keeps the user hooked and engaged with the platform. However, from the perspective of the user, the experience of being nudged into watching another show is not necessarily a positive one—especially if we take into account how their decision-making process is affected. In a recent article, Matthews [21] argues that the way Netflix influences their users reveals the true nature of the company's personalized environment "that acts as a set of blinders which constrain the agency of the audience through an interface designed to dazzle and disorient Netflix users." The author shows that key desires of users get obfuscated by an arbitrary presence or absence of certain browsing features, the existence of hidden menus, and the prominence of algorithmic recommendations. Siles et al. [28] describe Netflix as a heavily designed and engineered platform of "mutual domestication." This means that, on the one hand, users are incorporating algorithmic recommendations into many aspects of their everyday life (e.g., using the platform while they are doing other things, such as eating dinner or folding laundry) while the platform's goal, on the other hand, is to keep the users attached with the hope to turn them into ideal consumers. Matthews [21] states that the platform's design (e.g., graphics, menus, buttons, the catalog) and the algorithms used for giving recommendations are based on the continual tracking of user behavior and algorithmic data processes, and thereby comes to the conclusion that users ultimately keep "exchanging their agency with the designers of the platform through the algorithms for personalized content."

The agency-influencing relationship that we can see between users and algorithms with the example of Netflix today allows us to anticipate the impacts on users of future applications that might use even more sophisticated predictive algorithms. So far, recommendation systems are more or less restricted to the specific purpose of guiding the streaming, creation, or presentation of content. However, the more impactful individual companies become, and the more their products influence our decision-making in increasingly private aspects of our personal lives, the more important it is to be wary of how these algorithms work and affect our agency. Netflix is one example that shows how a technology that seems harmless in its primary goal has been successful in introducing algorithms into our lives in ways that shape our behavior without us fully recognizing those effects.

At the same time, it is crucial to note that those algorithms are not developed with the goal to harm people. To the contrary, their primary aim is to make the life of people easier in one way or the other. Streaming services such as Netflix are trying to facilitate a good viewing experience. Social media companies such as Facebook are aiming at improving digital communication between their consumers. While those goals necessarily align with the individual companies' aim to be profitable in a capitalistic society, this fact alone is not turning their products into ethically questionable ones. As a consequence, doing ethics about current developments in tech should not state that algorithms are at risk of usurping our agency simply, because big tech companies are working on them for monetary reasons. Rather, this ethical research should consist of a nuanced analysis that points out the ways in which AI technologies are overstepping certain agentive boundaries we identify as crucial to remain untouched.

There are two potential issues we need to face: First, the ordinary user might not perceive predictive algorithms (such as those employed by Netflix) as a threat to their agency. This lack of awareness urges an immediate assessment of the multiple ways in which future technologies are going to be embedded in our lives [3]. Second, it is important to recognize that technologies facilitated by predictive algorithms are constantly evolving with the potential to become more widespread and persuasive. In the future, the technology might not only track one specific aspect of our lives, but act as a constant observer and intervener that has the capacity to influence our everyday decision-making.

One of those technologies that might enter this personal space in the foreseeable future is augmented reality (AR). In general terms, AR is the superimposition of a computer generated image on real world imagery that is perceived by an observer [32]. While those projections can be made in a room with special equipment (such as video projectors),

[2] While many businesses struggled in 2020 due to the Covid-19 pandemic, Netflix increased their subscribers to 203.67 million [20].

there is a recent trend towards implementing this technology into head-mounted displays that use head tracking and depth-glasses to display computer generated images on the glasses the user is wearing [32]. Those AR glasses allow a user to have a multitude of functions displayed on objects that are perceived through those glasses in the real world—such as reading incoming text messages projected on a wall, getting a visual of the weather forecast on a bathroom mirror, or perceiving a picture of who is calling on a desk.

While AR technology is still in its early stages, a variety of researchers aim at testing the usability of AR in, among other areas, educational settings [12, 17], industrial maintenance [26], tourism [34], and surgery [32]. Parallel to this research being conducted, Facebook aims at providing a future in which a user can benefit from AR glasses to improve their everyday life. These glasses could accompany users during their entire day and offer recommendations through sophisticated predictive algorithms about actions they could perform next. In a recent blogpost, Facebook Reality Labs introduced such a functionality in their AR glasses as one of their major features: “the intelligent click” [10]. The intelligent click is an action prompt presented to the user that can be accepted, declined, or changed by microgestures (e.g., by tapping the index finger against the thumb). As Facebook illustrates on their website, the system might offer the prompt “Play running playlist?” if it detects that a user is heading outside for a jog and, based on past behavior, is most likely to want to listen to their running playlist. The algorithm that is producing those action prompts is generating the offered options based on the previous choices and the personal history of the user. FRL is introducing this technology as a crucial step for users to save time and to not get derailed from their “train of thought or flow of movement” [10]. Ultimately, as Facebook mentions in the video attached to their blog post, this technology is claimed to lead to “increased agency.”

However, as we have already seen with the example of Netflix that is using predictive algorithms to recommend more content to their users, eventually leading to mutual domestication between the user and the machine, predictive algorithms are not necessarily increasing user agency. Instead, they bind the user and the machine together in ways that we might not want or realize. This is why it is important to develop those technologies with a conceptual understanding of agency and (ideally) integrate that understanding in the design and development process. Facebook’s description of their future AR technology and the way it is going to impact their users indicates that they work with an overly simplistic understanding of agency in which our agentive capacities increase parallel to the ease in which our actions are carried out. From the perspective of FRL, this might include, but is not limited to, increasing the speed, scalibity, and number of options of interconnected systems. Given those complex ways in which algorithms can influence our agentive capacities, it is crucial to structurally analyze how predictive algorithms can be designed in ways that the agency of agents is maintained and not limited. Since this technology has the potential to impact thousands of lives in the foreseeable future, it is absolutely crucial to start this analysis now.

3 A conception of agency as a theoretical basis for the ethical analysis of predictive algorithms

When it comes to creating a future in which users might be aided by personalized predictive algorithms, it is important to recognize how user agency can be affected.[3] The following section offers a definition of agency that can be used as a conceptual background to develop these types of technologies in an ethically responsible way. For the scope of this paper, the definition will be tailored towards several functions in which agency can be realized in AI and, thereby, influence similar agentive capacities in human agents.[4]

An insightful take on this is proposed by Dignum [8]: 16 who understands agency as the “capacity to act independently and to make own free choices.” In their work, Floridi [11] and Dignum [8] further distinguish agency into the three characteristics of autonomy, adaptability, and interactivity. Autonomy denotes the capacity to act independently and to make own free choices. In this sense, human agents can be considered autonomous if they are not restricted by a system or, if anything, only aided in meaningful ways that ultimately increase their autonomy. Adaptability is the capability to learn from one's own experiences, sensory inputs, and reactions with others to react and adapt to changes in the environment. Human agents can be considered more adaptive if the decisions they make when confronted with sudden irregularities are not solely based on blindly following a system’s recommendation. Interactivity is the ability to perceive and interact with others. Human agents can be considered interactive if the way they create, maintain,

[3] There are many other ethical issues discussed in the literature such as algorithm privacy [9], bias [16], and trust [33]. While those issues are connected to agency in numerous direct and indirect ways, this paper will concentrate on the influence of AI on the agentive capacities of their users.

[4] It should be noted that the literature offers a variety of definitions, criteria, and viewpoints for human agency. In the philosophical subfield of action theory, agency is tied to intentionality of a person performing an action [4], [25]. While the three dimensions introduced in this paper are not sufficient to capture the whole phenomenon of agency, they are exceptionally well suited to point at those agentive capacities that might be taken over by an artificial system. This makes them an ideal candidate to analyze the agentive relationship between human agents and AI powered devices.

and end their relationships is not taken over by a system's algorithm. While this dimension of interactivity might seem somewhat similar to autonomy (i.e., being less influenced by AI), it differs insofar as it is not focusing on actions of a single actor but on the dynamics and interrelations between several human agents.

It is noteworthy that Dignum [8] introduces these three characteristics mainly as criteria of "AI agency", thereby indicating that they can also coincide with what we would understand as human agency. The reason for this framing is that current AI systems are developed with the goal to meet those three characteristics. As such, if AI technologies are developed in a way that they are autonomous, adaptive, and interactive, then they hold the status of agentive systems. Likewise, if humans meet those criteria, then they can be considered agentive humans. When agentive systems and agentive humans interact with each other, then it depends on their interrelation across those characteristics whether more agency is held by the system or by the human. As such, the human capacity to act autonomous, adaptive, or interactive can either be diminished or aided depending on how those agentive capacities are implemented in the artificial system.

Recognizing this interrelation, one of the commonly found demands in the literature is that maintaining agency depends on an appropriate design for the predictive power of AI in a way that it fosters but does not undermine human autonomy, adaptability and interactivity, which together are linked to broader conceptions of human dignity, self-determination, and autonomy [31]. In general terms, if AI is created as a supplement or replacement of human decision-making or judgment [3], then it might result in limiting a person's agency [11]. This is why many authors call for a balance between the decision-making power or agency we want to retain for ourselves and the decision-making power or agency we want to delegate to artificial agents or algorithms [11].

Taken together, predictive systems that give recommendations to their users should aid those characteristics of agency. While certain algorithms might support their users to a certain degree, those systems should not limit, restrict, or replace the agentive capacities of their users entirely. As such, it is crucial to find a balance between the amount of agentive capacity that is maintained by the individual human and the amount of agentive capacity that is taken over by the system.

4 The impact of predictive algorithms on agency: three case scenarios

While the previous section provided a conceptual understanding of the individual characteristics of agency, using those distinctions alone, it is unclear how they are going to play out. This technology is simply not available yet and, therefore, the scope in which it is going to impact human lives, is not immediately accessible. Nonetheless, it is important to start this ethical analysis of agentive capacities between agentive humans and agentive systems now. Otherwise, those technologies might enter our lives in a slow but steady manner until they are part of us in such complex and impactful ways that disentangling ourselves from them is extremely difficult [3]: 57).

One strategy to address those issues now is to identify sociotechnical contexts in which people and algorithms interact [14]. This contextualization can be done through thought experiments that offer a detailed scenario for identifying the ways in which AI technology can be agency increasing or agency decreasing. Against this background, the following section introduces hypothetical scenarios that use Facebook's intelligent click feature as an illustrative example to show how AI facilitated recommendation technologies could influence the three identified characteristics of agency. All scenarios are structured around a usage scenario indicated by Facebook according to which a user is assisted by an AR device while putting on their running shoes. Starting from this example, three hypothetical scenarios are introduced that match the three identified characteristics of agency, showing how such a system might be agency limiting. The following section considers how the technology could be changed to support human agency.

4.1 Hypothetical case scenario #1: a threat to autonomy

Autonomy is the first characteristic of agency and describes the capacity of agents to act independently and to make their own free choices. Regarding the relation between agentive humans and agentive systems, human autonomy is potentially limited if the decisions and actions of human agents are not regarded as their own. Here is a hypothetical scenario in which a user gets recommendations from an AI facilitated system that is intended to increase their autonomy but fails to do so:

> James is a 32 year old engineer who is tired of his lack of exercise during the COVID-19 pandemic. He decides to make the best of his situation and sets a goal to be more physically active. Deliberating among the options of bodily exercise available to him, he decides to go running once every three days. He is intrigued by Facebook's new assistive device and buys one of their AR headsets as an aid to stay motivated during his exercise endeavor. When he activates the headset for the first time, the system notifies him that the highest accuracy of prediction is achieved if it stays always on. Given that James has plenty of other devices that

> are always on, like his gaming console, his laptop, and his phone, this recommendation seems reasonable. He decides to wear his AR glasses not only for running but in a way that is similar to people who wear corrective glasses - basically throughout the whole day; from getting up in the morning to going to bed at night.
>
> After a week of using the system, the predictive algorithm has successfully identified James's behavioral patterns and starts recommending action items that are perceived as helpful to him to increase his agency in relation to running. On the one hand, the system takes over some actions, e.g. by keeping track of the music he listens to (e.g., "Play running playlist?") or by scheduling the next exercise ("Schedule next running exercise in two days?") while on the other hand, the system suggests actions for James to perform, e.g. by identifying or recommending exercises to do before he gets out (e.g., "Warm up with a stretching exercise?") or recommending food to eat after he is done (e.g., "Eat healthy snack in 30 minutes?").
>
> Little by little, the AI is going beyond putting out prompts concerning James' desire to exercise and begins to structurehis daily routine, habits, and preferences. What started off as a helpful reminder and personal aid to make things more convenient slowly turns into a streamlined recommendation system that is conditioning him to act upon certain actions. If James is not following those prompts, the system keeps nagging him until he gives in. This makes it increasingly difficult for him to come up with his own course of action. Over time, he feels constrained by the options recommended to him and perceives them as a diminishment of his autonomy.

As AI recommendation systems get introduced pervasively in our lives, they can be helpful for making decisions that are otherwise too difficult or complicated. However, at the same time, it is crucial to avoid scenarios in which our choices get overruled by technology. To understand what happened to James in this hypothetical scenario, we can ask the question in what ways our own choices are constrained if a system structures our weekly routine. The scenario depicted here is that the system ends up making decisions for the user by offering a rigid structure about what to do next, resulting in the problem that the AI takes the decision-making capacity away from the user.

The algorithm and the user standing in an intricate relation of autonomy to one another is a theme that can be commonly found in the literature. Susser [30] notes that it is, at least to some degree, inevitable that AI influences our perceived array of choices (choice architecture) and decisions. After all, Floridi et al. [11] argue that it is part of the concept of using AI that we give some of our decision-making power to the machines. However, Floridi and colleagues note that the delegation of autonomy should not fall on the algorithm entirely but rather should be protective of the intrinsic value of human choice. To develop a responsible algorithm that is not holding too much decision making power, Sundar [29] holds that AI should function as a decision aid but not as a decision maker. For instance, imagine James ignoring the recommendations to go for a run and going to the couch instead. If the device keeps being insistent and keeps bugging him ("Should I start the playlist to get you in the mood for running?") it might feel hard to resist. If that is the case, James might surrender his autonomy to the machine.

Constraining the decisions of a user like that can result in changes of a user's dispositions. In a number of experiments, Adomavicius et al. [1] show that the recommendations of a system primes the preferences of their users. Over time, they tend to take over the recommendations of a system in a way that they shape their disposition about what to do in the future. Priming creates habits and actions that might not be there with the decisional aids from the machine. Individuals might prefer not to have those habits and actions if given a more robust option or time to reflect. Given those considerations, users of recommendation systems are at risk of making a set of choices that are detached from their own preferences. Here, the AI became a decision maker that is their user's preferences. To find a solution to this problem, it is crucial to present choices in a way that they are not constraining but enabling the user to perform actions by developing systems that have a limited capacity of autonomy.

For a system like the FRL–AR glasses to be less autonomous, it is crucial to redirect their capacity to create decision-making recommendations (i.e., the device suggests a new activity) into a capacity to give decision-aiding recommendations (i.e., the device contributes to an already started activity). For instance, imagine a home of a user, where dirty clothing is lying on the floor. In this case, a decision-making recommendation that we are trying to avoid would consist in the system detecting the clothing on the floor and creating an action prompt based on that information (e.g., "Do you want help to do your laundry?"). Here, the problem consists in the system creating a new course of action that might not align with the user's current preferences.[5] To solve that problem, the system could instead offer a decision-aiding recommendation. For instance, imagine that the user is in the process of loading the washer. The system might use this information of the currently performed action to automatically detect the settings that are usually preferred by the user (e.g., temperature, wash cycle) and generate an appropriate decision aiding recommendation (e.g., "Start the washer with the settings

[5] The only exceptions are action prompts that are intentionally set by the user as a reminder to start the respective activity or offering more autonomy-preserving choices. In the former case, the user would be the decision-maker by setting up appropriate alarms, or allowing the

you like?”). Here, the system is aiding the user in a decision, because that decision was already made by that user and the system’s support is not interfering with the user’s overall preferences. One way to implement this feature is to not solely focus on the ways in which algorithms could (potentially) promote the autonomy of human agents but also to rigorously constrain the autonomy of algorithms [31].

4.2 Hypothetical case scenario #2: a threat to adaptability

Adaptability is the second characteristic of agency and describes the capacity of agents to learn from new experiences and to react to changes in the environment. Regarding the relation between agentive humans and agentive systems, adaptability is potentially limited if the agent is compromised in comprehending and acting upon those perceived changes. Here is a hypothetical scenario in which a user gets recommendations from an AI facilitated system that is intended to increase their adaptability but fails to do so:

> James enjoys running with his FRL-AR device. His favorite route is 3 miles long and leads through a narrow one-way street that eventually opens up to a secluded park. One day, however, his beloved secret park entrance is not accessible due to ongoing constructions blocking the entire street. James thinks about ways to get to this park on a different route, tries to visualize the neighboring streets, and starts running towards another park entrance he believes to be close. Shortly after he starts running, the FRL-AR glasses detect a deviation from his usual path and offer him another prompt: “Display the shortest route to your destination?”
> Since James was nervous to explore a different route on his own, he is happy about the convenience the recommendation offers him. He accepts the prompt but is puzzled when the system urges him to turn around and take a completely different route. After a little while of deliberation, James shrugs with his shoulders, quietly mumbling towards his AI glasses “You probably know better than me.”, and starts running according to the recommended path.

When we are in different environments or are experiencing a change on a previously known path, it can be helpful to rely on algorithms that help us out. Think about the ways in which the GPS on your phone helps you to find a new address or, if there is a sudden road blockage, quickly computes an alternative path to your destination. The difficulty for all kinds of tasks that are taken over by an algorithm is that the computation of recommendations can happen in ways that obscure the user’s ability to adaptively react to changes in the environment themselves. To understand what happened to James in this second hypothetical scenario, we can ask the question in what ways we can become over-reliant on algorithms for solving novel problems. The scenario depicted here is that the system is taking over the execution of a task.

To unravel the relationship between humans and machines in that scenario, it is important to think about the influence of algorithms on our capacity to make self-governed choices. Generally speaking, AI facilitated systems can be useful when they present us with a diverse set of choices to choose from. For predictive recommendation algorithms, the set of available options is tailored to the user based on information collected about individual preferences, aspirations, and vulnerabilities [30]. However, the way that this information is used and computed to come to an algorithmic decision can be entirely obscure to the user. And yet, if the level of trust is high enough, it can influence a user’s capacity to engage in active decision making. In a recent study, Logg et al. [19] found that people adhere more to advice when they believe that it is coming from an algorithm and not from a person. In those scenarios, users often place a greater level of trust in the output of the artificial systems than in other people or even their own knowledge, belief, or skill.

In the literature, there are several factors discussed that play a role on how the reliance of a user on a system builds over time. Susser [30] notes that one of the major reasons is the power of habituation. Using a device on a regular basis integrates it in our everyday decision making. The more we are used to a device aiding our adaptive decision making, the less we notice the influence it has on us. Boddington [3] explains that this tight embeddedness of algorithms in our lives makes those devices invisible to us, which in turn affects our ability to reconsider our course of action or to think about alternative choices. If technology becomes invisible like this, then the way that it structures and influences our decisions becomes invisible too, making us susceptible to manipulation [30]. Over time, this process can spiral into us “cognitively outsourcing” the task to the trusted machine, thereby limiting our own ability to be adaptive in new scenarios [29]. Independent of the concrete functional implementation, the danger consists in the user surrendering to algorithm-generated recommendations even if those recommendations are inferior [2]. This outsourcing of cognitive skills creates new dependencies between the user and the system. Over time, this can result in the user relying more

Footnote 5 (continued)

algorithm to be notified if certain criteria are met. In the latter case, in addition to a simple affirmation through clicking “yes”, action prompts could also be accompanied by other agency preserving prompts such as “no” or, if the system is constantly nagging, a “leave me alone” button).

and more on the output of the system while existing skills to solve the original problem either wither or fail to develop in the first place. If a certain skill is always taken over by a machine, then that skill will erode over time.

A good example of skill erosion through AI is the way in which GPS is integrated into our lives for navigation. In modern day society, people use their own navigational skills mostly on rudimentary levels and instead rely on the AI system for computing and displaying their routes. In those cases, users often put too much trust into their devices. Johnson et al. [18] state that such an over confidence in the system results in users failing to notice the occurrence of faults and errors. They further show that blindly following the recommendation of an AI system for navigation is a major cause for accidents, delays, and traffic. One specific example that shows the impact of following GPS routes without reflection is the way in which the majority of overpass accidents in the State of New York are caused by incorrectly working algorithms for navigation [27]. In those cases, the drivers put such a high amount of trust into their navigational system that they stop paying attention to their surrounding environment, such as road signs that indicate the height of bridges. Here, the algorithm interferes with the user's senses to perceive and act upon changes in the environment.

GPS tracking and its integration into predictive recommendation algorithms is just one example that shows how users can overtrust AI in ways that their skills derode and their confidence in making informed decisions based on those skills in a changing environment gets negatively impacted. As another example, imagine driving safely through town while the weather suddenly shifts from sunny to snowy. An AI assisted driver might never learn how to adapt their driving to snowy conditions if the AI controls the vehicle through that transition.[6] With the algorithms used in the AR–FRL system, there are numerous other ways in which users might blindly follow the system's recommendations due to their high trust and reliance on the system's outputs.

One solution to the issue of over-reliance on tasks or skills that are taken over by predictive recommendation algorithms consists in making the algorithms more transparent. For instance, if James is turning to the system to assist him by computing an alternative path, then he should get a summary of the criteria that were used to compute that alternative route. This allows him to take part in the skill of navigating, since he can help identify certain determinants that are important to him, change them if needed, and engage in a more direct form of human machine interaction. The algorithms can also be designed in a way that allows the user to share their preferences about certain scenarios that may occur. For instance, James could define upfront that an alternative route should be computed according to set parameters, such as well-maintained roads, sidewalks, minimal hills, elevation gain, or scenic views. During the computation of a new route in a new environment, those criteria can be fed back to him, thereby allowing him to make an informed decision based on the information transparently unfolding in front of him. Another option consists of giving James several alternative routes with openly communicated criteria. He can then choose among a variety of pre-selected routes and choose the one he sees fit the most. For all those different scenarios, his own skill of navigating through a new environment will be taken over by some degree. However, this sort of transparent interaction with the device does happen according to his terms and is likely to result in the development of a new skill on how to most effectively use the system to adaptively react to novel situations.

4.3 Hypothetical case scenario #3: a threat to interactivity

Interactivity is the third characteristic of agency and describes the capacity of agents to perceive and interact with others. Regarding the relation between agentive humans and agentive systems, interactivity is potentially limited if the relation to others is dictated or influenced by the system. Here is a hypothetical scenario in which a user gets recommendations from an AI facilitated system that is intended to increase their interactivity but fails to do so:

> James keeps running on a regular basis and makes a continued effort of doing exercises. However, he feels lonely on his daily runs, so he enlists the AR glasses to join a Facebook running group in his area. When he is going to their usual meeting place, he is welcomed by friendly strangers and starts exercising with them. During their mutual run, he starts chatting with Lydia who is around his age and also started running recently with FRL-AR glasses as a motivator. Over the next weeks, they continue to see each other in the group, often running next to each other. Recognizing their physical closeness over time, their respective FRL-AR systems recommend each other as friends, which they both accept.
>
> In the days that follow, James would love to meet Lydia for a coffee but, at the same time, is not willing to put a lot of effort into taking their relationship to the next level. Due to his lack of engagement, he asks the FRL-AR system for dating advice. In what follows, the algorithm offers him real time tips about Lydia's preferences. For instance, the system remembers what kinds

[6] I want to thank my anonymous reviewer for suggesting this example.

> of movies she likes when he has forgotten that they even had this conversation. Based on their recorded interaction data, the system recommends James to ask Lydia out for her favorite type of movie ("Ask Lydia to watch a horror movie?"). He accepts the prompt which results in an automated message sent to Lydia through the FRL-AR app ("Do you want to watch a horror movie together anytime soon?"). Lydia assumes that the message is genuine, feels understood, and accepts. In the conversations that follow, James does not want to put effort into remembering all that information and is relying more and more on the system's recommendations in order to find topics that keep Lydia engaged. She is not aware of that and believes that she found someone who truly gets her.

Predictive recommendation algorithms can influence how we form relationships with others. Recommendations about new friends or people we meet for the first time while being in an unfamiliar social setting can be helpful to keep track of new people and to enlarge the circle of acquaintances. However, there are also limits to what a recommendation system should be able to do. While the interaction with Lydia seems "successful" in some sense, the way how that conversation plays out might be seen as not genuine or even misleading to Lydia. Furthermore, it could be argued that she is robbed of her agency to make an informed decision about her interaction with James. To understand how the algorithm went too far, we can ask the question how genuine our interactions with friends are if they are based on data suggested by an algorithm. The scenario depicted here is that the system is shaping how people interact, leading to the problem that the genuineness of their interaction can be questioned; and thereby their ability to make own decisions based on that perceived genuiness.

People are often seeking the help of algorithms to meet new people; for extending their professional network, seeking other people to spend time with after work, or finding a potential partner through dating apps. With the rise of this technology, it is crucial to remember that it is not just two people who shape their relationship as equals but that there is also a software involved that encodes values and decisions about what is deemed as important, leading to what Bucher [5]: 490f.) coined "programmed sociality." Facebook is known to shape friendships within their platform by heavily relying on algorithms. As Chambers [6] illustrates in her paper, this "algorithmically engineered friendship" can lead to changes in public intimacy, privacy, and trust.

While those are impactful ethical side-effects of technologies we already see today, in the future, those issues might get exacerbated when new forms of communication arise that rely more on algorithms. In their recent work, Hancock et al. [15] coin the term "Artificial Intelligence-Mediated Communication" (AI-MC) which denotes not only the ways in which AI shapes the friendships we acquire but how algorithms operate on behalf of the user to modify, augment, or generate messages people send to each other. The authors state that this is likely to influence both the sender on how to present themselves and the receiver on whether the communication is perceived as trustworthy or authentic. Hancock and colleagues add that this will also have consequences on how interpersonal dynamics are shaped through self-representation, impression formation, and trust.

The problem of getting real time recommendations is that it takes away a core value of creating and maintaining relationships: genuineness. In the hypothetical example depicted above, Lydia is not only continuing the conversation due to the mere fact that James wants to see horror movies, but because she is under the impression that he actually listened to her and genuinely cares about her interests. However, if AI recommendation systems are powerful enough to shape the communication of people in a way that the authenticity of their messages can be questioned, then they are at risk of replacing a genuine exchange of love and care with probabilistic judgments about topics of interests that lack the sincere characteristics that are at the core of meaningful social interactions.

This potential risk in AI mediated social interactions is further exacerbated by the fact that AR is likely to generate new ways of communication that go far beyond sending and receiving text messages. For instance, AR glasses might be used to project other people into the individually perceived environment, such as a virtual image of a friend sitting on a chair at your kitchen table. While this change in format opens new forms of immersive communication that can be beneficial, Miller et al. [23] have already shown that it also negatively influences people's task performance, nonverbal behavior, and social connectedness, especially in larger group gatherings, where only some but not all people have access to that technology.

As of now, it is not clear how much those novel social features that might be possible in the future are going to shape the interactivity of people. Apart from their overall feasibility, their impact also depends on what Chambers [6] calls the "scale of sociality", i.e., what type of conversations are going to be typical with what kind of device. She illustrates that term by showing how social media apps are used for different purposes. For instance, WhatsApp has group limits for up to 20 people and is mostly used for private and intimate connections while the group scales on Facebook and Twitter reach thousands of people, thereby influencing the content shared and the type of conversations held. If the algorithms for AR are embedded with social media apps, then they have immense power to shape how people generate new friendships, maintain existing ones, and communicate with each other—and what it means to be genuine in those interactions.

Overall, there are numerous ways in which algorithms affect the interactivity of people who are trying to get to know each other or engage in a conversation. Going back to the example of how Lydia and James met, there are several ways in which their agency is potentially impacted. Using a device as a means for interaction is not necessarily problematic but having algorithms that drive the content of a conversation might result in building unwarrented trust—like in the case of Lydia who is lead to believe that James is genuinely interested in her. There are many questions in this scenario alone that are in need of a more throrough examination: How do we design, build, and advertise this technology in a way that it preservers people's interactive capacities and their agency? How can we support genuine interactions without becoming too intrusive on intimacy? It is crucial to actively pursue those and other follow-up questions while the technology is advancing to offer oversight that aims at protecting human agency as much as possible.

5 Conclusions

There is an overwhelming amount of research—a lot of it done in industry—that aims at developing future AI systems which are likely to impact the lives of hundreds of thousands of individuals. Many advances those technologies are expected to bring have a reasonable chance of being supportive and helpful for improving people's lives. However, they can also have negative consequences that can diminish the agentive capabilities of their users.

This paper focused on the predictive recommendation algorithm of the "intelligent click" that is currently in development by Facebook Reality Labs for their future AR headset. The goal was to illustrate how the underlying algorithm might shape, influence, and redirect the agentive capacities of people using that device. To generate a conceptual foundation of those agentive capacities, agency was defined by the three dimensions of autonomy (i.e., acting independently), adaptability (i.e., reacting to changes in the environment), and interactivity (i.e., interacting with others). Hypothetical thought experiments were introduced that point at the ways in which the AI system could take over those agentive capacities in unwarranted ways. To redirect the design of such systems, a variety of relevant questions were asked that revealed what types of human machine interactions should be avoided in the future.

For understanding autonomy, we can ask the question in what ways our own choices are constrained if a system structures our weekly routine. The scenario depicted here is a system that ends up making decisions for the user. For ensuring human autonomy in future predictive recommendation systems, we must ask how we can offer decision-aiding recommendations without prompting decision-making recommendations. For understanding adaptability, we can ask the question in what ways we can become over-reliant on algorithms for solving novel problems. The scenario depicted here is a system that is taking over the execution of a task. For ensuring human adaptability in future predictive recommendation systems, we must ask how we can provide relevant information without presuming preferred adaptations or taking over a skill entirely. For understanding interactivity, we can ask how genuine our interactions with friends are if they are based on data suggested by an algorithm. The scenario depicted here is a system that is shaping how people interact. For ensuring human interactivity in future predictive recommendation systems, we must ask how to support genuine interactions without becoming too intrusive on intimacy.

To mitigate those ethical issues for future AR products, it is imperative to implement an analysis of the user's agentive capacities alongside the early stages of the technology. This analysis can serve as a foundation for the ethical oversight during the design and development processes of emerging technologies in ways that human agency is protected at all times. Only if those issues are taken seriously and acted upon immediately, predictive recommendation algorithms such as the intelligent click can be developed in responsible ways—instead of being recognized too late when the technology is already on the market and widely used.

Funding This work was supported by the National Science Foundation (EEC-1028725).

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

References

1. Adomavicius, G., Bockstedt, J.C., Curley, S.P., Zhang, J.: Do recommender systems manipulate consumer preferences? A study of anchoring effects. Inf. Syst. Res. **24**, 956–975 (2013)
2. Banker, S., Khetani, S.: Algorithm overdependence: how the use of algorithmic recommendation systems can increase risks to consumer well-being. J. Public Policy Mark. **38**, 500–515 (2019)
3. Boddington, P.: Towards a code of ethics for artificial intelligence research. Springer, Berlin Heidelberg, New York (2017)
4. Bratman, M.: Intention, plans, and practical reason. Harvard University Press, Cambridge (1987)
5. Bucher, T.: The friendship assemblage: investigating programmed sociality on Facebook. Television & New Media **14**, 479–493 (2013)
6. Chambers, D.: Networked intimacy: algorithmic friendship and scalable sociality. Eur. J. Commun. **32**, 26–36 (2017)

7. Crews, C., Colson, C., Elson, R.: It does matter who your friends are: a case study of Netflix and "friends" licensing. Global J. Bus. Pedagogy **4**, 6–13 (2020)
8. Dignum, V.: Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer International Publishing, Cham (2019)
9. Dilmaghani, S., Brust, M.R., Danoy, G., Cassagnes, N., Pecero, J., Bouvry, P.: Privacy and Security of Big Data in AI Systems: A Research and Standards Perspective. In: 2019 IEEE International Conference on Big Data (Big Data). Los Angeles, CA, USA: IEEE. 5737–5743. Online available: https://ieeexplore.ieee.org/document/9006283/ (2019)
10. Facebook Reality Labs: Inside Facebook Reality Labs: Wrist-based interaction for the next computing platform. In: Tech@Facebook. Online available: https://tech.fb.com/inside-facebook-reality-labs-wrist-based-interaction-for-the-next-computing-platform/ (2021)
11. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E.: AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Mind. Mach. **28**, 689–707 (2018)
12. Garzón, J., Pavón, J., Baldiris, S.: Systematic review and meta-analysis of augmented reality in educational settings. Virtual Reality **23**, 447–459 (2019)
13. Gomez-Uribe, C.A., Hunt, N.: The Netflix recommender system: algorithms, business value, and innovation. ACM Trans. Manag. Inf. Syst. **6**, 1–19 (2016)
14. Green, B., Chen, Y.: The principles and limits of algorithm-in-the-loop decision making. Proc. ACM Hum-Comput. Interact. **3**, 1–24 (2019)
15. Hancock, J.T., Naaman, M., Levy, K.: AI-mediated communication: definition, research agenda, and ethical considerations. J. Comput.-Mediat. Commun. **25**, 89–100 (2020)
16. Harris, C.: Mitigating cognitive biases in machine learning algorithms for decision making. In: Companion Proceedings of the Web Conference 2020. Taipei Taiwan: ACM. S. 775–781. Available: https://dl.acm.org/doi/https://doi.org/10.1145/3366424.3383562 (2020)
17. Ibáñez, M.-B., Delgado-Kloos, C.: Augmented reality for STEM learning: a systematic review. Comput. Educ. **123**, 109–123 (2018)
18. Johnson, C.W., Shea, C., Holloway, C.M.: The role of trust and interaction in GPS related accidents: a human factors safety assessment of the global positioning system (GPS). Vancouver (2008)
19. Logg, J.M., Minson, J.A., Moore, D.A.: Algorithm appreciation: people prefer algorithmic to human judgment. Organ. Behav. Hum. Decis. Process. **151**, 90–103 (2019)
20. Lozic, J.: Financial analysis of Netflix platform at the time of Covid 19 pandemic. In: Economic and social development. Rabat (2021)
21. Matthews, J.: Netflix and the design of the audience. MedieKultur **69**, 52–70 (2020)
22. Meta: Introducing Meta: A Social Technology Company. Online available: https://about.fb.com/news/2021/10/facebook-company-is-now-meta/ (2021)
23. Miller, M.R., Jun, H., Herrera, F., Villa, Y., Jacob, W., Greg, B., Jeremy, N.: Social interaction in augmented reality. PLoS ONE **14**(5), 2016290 (2019)
24. Newton, C.: Mark in the metaverse. Facebook's CEO on why the social network is becoming 'a metaverse company'. In: The Verge (2021)
25. Pacherie, E.: The phenomenology of action: a conceptual framework. Cognition **107**, 179–217 (2008)
26. Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. Robot. Comput –Integr. Manuf. **49**, 215–228 (2018)
27. Robbins, J.: GPS navigation&#x2026; but what is it doing to us? In: 2010 IEEE International Symposium on Technology and Society. Wollongong, Australia: IEEE. 309–318. Online available: http://ieeexplore.ieee.org/document/5514623/ (2010)
28. Siles, I., Espinoza-Rojas, J., Naranjo, A., Tristán, M. F.: The mutual domestication of users and algorithmic recommendations on Netflix. In: Communication, Culture and Critique (2019)
29. Sundar, S.S.: Rise of Machine agency: a framework for studying the psychology of human–AI interaction (HAII). J. Comput.-Mediat. Commun. **25**, 74–88 (2020)
30. Susser, D.: Invisible Influence: Artificial Intelligence and the Ethics of Adaptive Choice Architectures. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. Honolulu HI USA: ACM. 403–408. Online available: https://dl.acm.org/doi/https://doi.org/10.1145/3306618.3314286 (2019)
31. Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., Floridi, L.: The ethics of algorithms: key problems and solutions. AI & Soc. (2021). https://doi.org/10.2139/ssrn.3662302
32. Vávra, P., Roman, J., Zonča, P., Ihnát, P., Němec, M., Kumar, J., Habib, N., El-Gendi, A.: Recent development of augmented reality in surgery: a review. J. Healthcare Eng. **2017**, 1–9 (2017)
33. Winfield, A.F.T., Jirotka, M.: Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **376**, 20180085 (2018)
34. Yung, R., Khoo-Lattimore, C.: New realities: a systematic literature review on virtual reality and augmented reality in tourism research. Curr. Issue Tour. **22**, 2056–2081 (2019)