

# “DARK PARTNERS”: TRANSPARENCY OBLIGATIONS AGAINST DECEPTION IN VIRTUAL INFLUENCER MARKETING

Jacopo Ciani Sciolla

Università degli Studi di Torino (Italy)

[jacopo.cianisciolla@unito.it](mailto:jacopo.cianisciolla@unito.it)

## ABSTRACT

The paper examines the creation of realistic and visually appealing virtual influencers that take the form of hyper-realistic characters who can be nearly impossible to distinguish from real-life influencers. Since only a few national regulators provide a duty to disclose the influencer’s real nature, consumers may falsely believe they are engaged in communications with humans. This information is valuable for consumers who are more likely to rely on recommendations from individuals with views and beliefs similar to their own. Based on Habermas’ theory of communicative action and Kantian ethics, the overall aim is to suggest that brands must be transparent and should not engage in marketing communications referring to any virtual testimonial of products because such practice would not be based upon any bona fide use, nor personal opinions, beliefs, or experiences of human fellows.

**KEYWORDS:** digital marketing, dark patterns, virtual influencer, unfair commercial practices, transparency, ethics.

## 1. INTRODUCTION

The online environment is populated by internet companies exploiting users’ psychological vulnerabilities thanks to the use of AI. The aim is to maximise profits by nudging or deceiving users into making decisions that, if fully informed, they might not make.

Such practices normally fall under the “dark patterns” umbrella term, which refers to manipulative interface design choices that negatively impact the user’s decision-making, leading users to act against their interests,

The distinctive feature of dark patterns is the practice of user experience (UX) and user interface (UI) design to deceive users into accepting either unwanted purchases or subscriptions that depend on increasing levels of anxiety due to time limits and social pressure. UX and UI are conceptual design disciplines that revolve around the interaction between users and machines to shape systems and computer interfaces that address the user’s experience when using a platform (Dove et al. 2017). Good UX aims to provide people with interactions that are seamless, enjoyable, and intuitive. However, UX is a tool that can be used both for good and for evil. Dark patterns are one such category of evil design and given their growing use and the ease with which they can be added to platforms (i.e., dark patterns as a service), the focus is increasingly on the understanding of these practices, consequent harms, and potential countermeasures (EU Commission 2022; OECD 2022; BEUC 2022).

Dark patterns come in many different shapes, may employ different kinds of design-based elements, and can intervene at different stages of a transaction, such as the advertising one. New forms of dark patterns are constantly emerging, with new technologies and new kinds of user interfaces. Various

regulatory measures to respond to dark patterns have been proposed or implemented and calls among the UI/UX design community to adopt ethical standards have increased<sup>1</sup>.

This paper aims to study a phenomenon that has the same distinctive character of dark patterns and a very similar influence on end-users from a legal and socio-ethical perspective, but, until now, has not received the same attention by policymakers: the creation of realistic and visually appealing virtual influencers making consumers falsely believe that they are engaged in communications with humans. I refer to them as “dark partners” because they partner in commercial communications with brands, advocating for their products or services, without being transparent about their virtual identity.

Accordingly, the paper is structured as follows. Next, the focus is on the virtual influencer concept and some basic distinctions with real-life ones. The section illustrates why brands are increasingly using them. Section 3 highlights the risk that consumers may confuse virtual influencers with human beings and clarifies why this could be relevant for the success of an ad strategy. Section 4 analyses current law and stresses that only a few national regulators have established a duty to disclose the influencer’s virtual nature. Section 5 draws on Habermas’ theory of communicative action and Kantian ethics to assess this practice. The conclusions provide recommendations for policymakers.

## **2. VIRTUAL INFLUENCER MARKETING: THE REASONS FOR A SUCCESSFUL STORY**

Influencer marketing involves the promotion of specific brands or products through influencers using the positive impact they are likely to have on consumer perceptions.

There is no legal definition of an “influencer” enshrined in EU law. However, the UCPD guidelines (EU Commission 2021) specify that “an influencer is generally described as a natural person or virtual entity who has a greater than average reach in a relevant platform” (sec. 4.2.6).

Influencers are usually known for being experts in particular topics (e.g., travel, lifestyle) and for creating content carrying different values for consumers (Audrezet et al. 2020), including educational, entertainment, or advertising, often simultaneously. In addition to their content-making skills, influencers’ popularity is further driven by their ability to build strong connections and trust with consumers by revealing personal information that consumers can relate to (Penttinen et al. 2022); in a word, by making their life ‘transparent’ or ‘public’.

Traditionally, brands collaborate with real-life influencers (i.e., humans living in a physical world) who can make their own decisions regarding sponsored collaborations with brands and form opinions about the products and services they promote.

With recent technological developments, brands increasingly started to work with virtual influencers.

### **2.1. The notion of “virtual influencer”**

Virtual influencers are non-human digitally created characters sharing social media content and engaging in interactive communications to obtain influential status among consumers. Within this wide category, experts distinguish between those created with computer-generated imagery technology (CGI influencers) and AI influencers that rely on AI technologies in creating content and interacting with consumers.

Virtual influencers can have different forms. Some authors developed a taxonomy based on their similarity to human appearance, also known as anthropomorphism (Mende et al. 2019), and their placement on the reality-virtuality continuum (Hudson et al. 2019), ranging from unimaginable

---

<sup>1</sup> <https://www.design.org.au/code-of-ethics/dia-code-of-ethics>.

characters to hyper-realistic characters that can be nearly impossible to distinguish from humans (Mouritzen et al. 2023). Like real-life influencers, hyper-realistic human virtual influencers share content about their personal and social lives, which often features them in the physical world performing human tasks, including attending fashion shows and commercial photoshoots. Consider Lil Miquela who claims to be a 19-year-old AI robot with a passion for social justice, fashion, music, and friendship. Currently, Miquela has over 190,000 monthly listeners on Spotify and gives interviews at major events (Savageaux 2022). She has been featured in campaigns by Calvin Klein and Prada.

Accordingly, virtual influencer marketing can also be classified as mixed reality because it allows to mix objects from both physical and virtual worlds and makes the boundaries between the real and virtual world blurred.

It is worth noting that some digital characters appear to exist only in a virtual world, while some others are avatars of real-life celebrities (Arsenyan-Mirowska 2021). In September 2023, Meta launched 28 AI-powered chatbots featuring Kendall Jenner (Billie), Paris Hilton (Amber), and Snoop Dogg (Dungeon Master). Currently, they are only available for testing in the US but AI shall make celebrities, shortly, omnipresent, since they can penetrate every market and format at any time. Most of the time AI clones of celebrities, grabbing user’s attention on YouTube, are just scams relying on AI voice cloning paired with decontextualized video of the celebrity<sup>2</sup>.

Even if the language in advertising distinguishes between influencers, endorsers, celebrities, or ambassadors, for the sake of simplicity, I will use the term virtual influencers to refer in general to digital characters that represent brands in digital advertising, irrespective of the degree of reputation they enjoy with the public. Therefore, I will leave aside from my analysis virtual assistants and chatbots acting on the brand’s website to provide consumers support: they are more readily recognisable as such.

## **2.2. Unique features and advantages**

Similar to real-life ones, virtual influencers can establish relationships and engage with a large number of consumers on social media (Hugh et al. 2022). Customization allows for designing virtual influencers with attributes that appeal to specific target consumers or fit the values and image of the promoted brands (Conti et al. 2022). Notably, their visual appearances and behaviours can be modified following changes in market trends and evolving consumer preferences.

Marketing collaborations with virtual influencers are also likely to require less time and financial resources in comparison to working with humans (Arsenyan-Mirowska 2021). Because of their digital nature, virtual influencers do not have physical constraints and can be anywhere, anytime (Conti et al. 2022). This means that consumers can potentially interact with the same virtual influencer across different digital platforms simultaneously.

A key issue is that they are not subject to reputational damage. Just in case, they can simply be deleted, and companies can create a new one. Thus, in comparison to working with real-life influencers, collaborations with virtual influencers assume lower risks related to involvement in scandals and unethical behaviours (Guthrie 2020).

---

<sup>2</sup> [https://www.404media.co/joe-rogan-taylor-swift-andrew-tate-ai-deepfake-youtube-medicare-ads/?mc\\_cid=f250e2b063&mc\\_eid=f720a42bfb](https://www.404media.co/joe-rogan-taylor-swift-andrew-tate-ai-deepfake-youtube-medicare-ads/?mc_cid=f250e2b063&mc_eid=f720a42bfb).

### 2.3. The social response theory

Even though virtual influencers do not exist in real life, several studies show they are perceived as authentic, regarding their physical appearance, personality and behaviour (Moustakas et al. 2020). This is coherent with the social response theory (Festinger 1954), according to which when consumers come across virtual influencers, they engage with them as they do with real-life ones, by applying the same social rules of interactions with humans, though they know virtual influencers are not humans (Moon 2003). Hence, as long as consumers will respond to virtual influencers as they do to real-life ones, it is not surprising that the former are capable of being preferred to humans.

## 3. RISKS OF CONSUMER DECEPTION AND MISINFORMATION

Virtual influencers may have many attractive features for brands, but they also raise some concerns.

### 3.1. Distinguishing virtual from real-life influencers

As virtual influencers are designed to have human-like features and behaviours, these brand's commercial partners might be particularly difficult to distinguish from real-life ones. Much like deepfakes, the rise of "dark partners" highlights our inability to distinguish reality from fabrications.

Many warn of the serious consequences when we can no longer trust (Durante 2011) any of the information we consume. The prevalence of fake presences may eradicate our sense of reality in the virtual realm. AI-powered virtual influencers may be purposefully designed for or tricked into (e.g., by untruthful or low-quality online data) spreading misinformation and other unethical communications (Mustak et al. 2022). The background stories of virtual influencers, the content they share, and, most importantly, their visual appearance can create false representations in society, like unrealistic perceptions of beauty standards (Gill 2023). This can be problematic, i.e., consumers having difficulties distinguishing virtual from human influencers (Franke et al. 2023), as these consumers do not realize that they are comparing themselves to a non-human and may feel anxious about the way they look, to the point of inhibiting their ability to live well (Deng-Jiang 2023).

This risk gets worse when influencers are involved in marketing activities. As consumers are more likely to rely on recommendations from individuals that have views and beliefs similar to their own, making consumers falsely believe they are engaged in communications with humans, might suspend consumer's abilities to identify and critically evaluate persuasive marketing tactics.

### 3.2. The commercial risk of disclosing virtuality

Some virtual influencers are transparent about their virtual identity. However, this is not always the case. The reason is that disclosing it may negatively affect the effectiveness of the communication.

The source credibility model shows that influencers' perceived characteristics may impact their trustworthiness, expertise, and attractiveness and affect the desired results of their messages (Ohanian 1991). Having low source credibility, influencers will lose the ability to engage consumers with sponsored posts. In this regard, knowing exactly the human or non-human nature of an influencer is pretty relevant. The outcomes of marketing research show that anthropomorphism increases brand liking and purchase intentions, while disclosing virtuality may lead people to feel uncomfortable or become more suspicious of persuasion attempts (Woodroof et al. 2020).

Thus, disclosing virtuality in commercial communications may lead to lowered brand trust and attitudes, lower purchase intentions (van Reijmersdal et al. 2016), and engagement (Boerman 2020).

#### **4. A REVIEW OF THE LEGAL FRAMEWORK**

Currently, there are very few legislations providing guidance on what brands should do to avoid misleading consumers about the real nature of influencers. While there has been some initial research on virtual influencers in law and ethics, such studies are largely descriptive, mostly documenting the existence of these practices. The main objective of this section is to address this gap and design a legal and ethical benchmark to support a clear understanding of the lawful or unlawful nature of virtual influencers' marketing.

##### **4.1. Jurisdictions specifically addressing the virtual influencers practice.**

The first and, at the time of writing, only jurisdiction establishing that virtual influencers “must additionally disclose consumers that they are not interacting with a real human being” is India. Following the Consumer Protection Act of 2019, the Advertising Standards Council of India (ASCI) in 2021, became the first national regulator to require an “upfront and prominent” disclosure of this kind (Patnaik 2021)<sup>3</sup>.

India has been followed by France. The Influencers Act, which came into effect on 1 June 2023, supplements the pre-existing regulations on advertising establishing that content with altered or artificially intelligent images must be accompanied by statements such as “virtual images” in order to limit the psychological impact on the public.

In the U.S., the Federal Trade Commission released an updated version of the Endorsement Guides<sup>4</sup>, which makes clear that brands may be held liable for virtual influencer's unfair commercial practices as it happens for human endorsers. This means that virtual influencers should avoid making statements implying their humanity or a personal experience with the product.

Notwithstanding that, the Guides do not provide for any duty of disclosing virtual identity and some scholars have already recommended to fill this gap (Masteralexis 2021).

It is also noteworthy to say that the state of California recently introduced a ban from using avatars in political communication, but this is valid only “within 60 days of an election”<sup>5</sup>.

##### **4.2. Virtuous examples of self-regulation**

Platforms are also taking matters into their own hands. For example, TikTok updated its platform guidelines to require that synthetic or manipulated media that shows realistic scenes be clearly disclosed. This can be done using a sticker or caption, such as “synthetic”, “fake”, “not real”, or “altered”. The guidelines require disclosure to be directly in the videos, not just in the virtual influencer's bio<sup>6</sup>.

##### **4.3. Omission of material information under the Unfair Commercial Practices Directive**

Obligations to disclose virtual identity do not exist at the EU level. However, any BtoC practice that materially distorts or is likely to distort the economic behaviour of an average consumer normally amounts to a misleading practice, as regulated by the Directive 2005/29/EC (Art. 6) concerning unfair

---

<sup>3</sup> ASCI (2021). The Code for self-regulation of advertising content in India. See Article 1.4 of the Guidelines for influencer advertising in digital media providing that “A virtual influencer must additionally disclose to consumers that they are not interacting with a real human being. This disclosure must be upfront and prominent”.

<sup>4</sup> 16 CFR Part 255: Guides Concerning the Use of Endorsements and Testimonials in Advertising.

<sup>5</sup> A.B. 730, 2019 Leg., Reg. Sess. (Cal. 2019).

<sup>6</sup> <https://www.tiktok.com/creators/creator-portal/en-us/community-guidelines-and-safety/ai-generated-content-label/>

commercial practices (UCPD). Misleading practices could be either by actions or by omissions. Articles 7(1) and (2) establish a positive obligation on traders to provide all the ‘material information’ that the average consumer needs to make an informed purchasing decision.

The UCPD does not define ‘material information’. However, by way of interpretation, it is possible to argue that the virtual nature of an endorser should be considered as such.

Relevant to this purpose is Article 7.4 which mandates to disclose “whether the third party offering the products is a trader or not” even if limited to the specific case of an ‘invitation to purchase’ on online marketplaces. Furthermore, in the *Wathelet* case<sup>7</sup>, the Court stressed that “it is essential that consumers are aware of the identity of the seller”. The same principle should be extended to an influencer as a person ‘acting in the name of or on behalf of a trader’. Indeed, the EU Commission clarified that for the purposes of the UCPD, an influencer may be qualified as a ‘trader’ (EU Commission 2021). Consequently, the obligation to be clear about the identity concerns directly all persons that carry out promotional activities towards consumers on behalf of a trader.

#### **4.4. The duty to disclose the commercial intent of a commercial practice**

Other arguments in favour of a disclosure duty may be derived by analogy with other information requirements established directly by EU law<sup>8</sup>.

First, Article 6(a) of the e-Commerce Directive<sup>9</sup>, Articles 9, 10 and 28(b) of the Audiovisual Media Services Directive<sup>10</sup>, similarly to Article 7(2) UCPD, establish that failing to identify the commercial intent of a practice is regarded as a misleading omission.

Coherently, the driving force behind all influencer marketing regulation adopted by advertising self-regulation authorities is the principle that influencers must disclose when they have a material connection with brands they promote through clear and understandable disclaimers such as #ad or #sponsored (Ciani-Tavella 2017).

Second, the UCPD prohibits as misleading by default “falsely claiming or creating the impression that the trader is not acting for purposes related to his trade, business, craft or profession or falsely representing oneself as a consumer”.

The purpose of this information requirement is to make sure that consumers always understand the very nature of the communication and know with whom they are interacting online. The same transparency interest exists when consumers are facing virtual influencers.

#### **4.5. Transparency obligations under the Consumer Rights Directive and the Digital Services Act**

EU law is generally averse to any form of hidden marketing.

Article 8(5) of the Consumer Rights Directive<sup>11</sup>, in the case of telemarketing, held the trader to “disclose the identity and, where applicable, the identity of the person on whose behalf he makes that call, and the commercial purpose of the call” (Luzak 2015).

<sup>7</sup> EUCJ, 9 November 2016, *Sabrina Wathelet v Garage Bietheres & Fils SPRL*, C-149/15, para 37.

<sup>8</sup> Article 7(5) UCPD clarifies that ‘information requirements established by EU law in relation to commercial communication, including advertising’, shall be regarded as material information by “default”.

<sup>9</sup> Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market.

<sup>10</sup> Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation, or administrative action in Member States concerning the provision of audiovisual media services.

<sup>11</sup> Directive 2011/83/EU on consumer rights.

In the same direction, social media platforms have recently seen their transparency obligations reinforced by the Digital Services Act<sup>12</sup> (Cauffman-Goanta 2021). Recital 68 states that “providers of online platforms should...be required to ensure that the recipients of the service have certain individualised information necessary for them to understand when and on whose behalf the advertisement is presented”. Article 26 establishes that for any advertisement presented, the recipients should be “able to identify, in a clear, concise and unambiguous manner and in real time, the following:...(b) the natural or legal person on whose behalf the advertisement is presented; (c) the natural or legal person who paid for the advertisement if that person is different from the natural or legal person referred to in point (b)”<sup>13</sup>.

Based on these principles, it is not surprising that the European Consumers Organisation (BEUC 2023) recently supported the introduction of two disclosure obligations regarding “edited” or “altered” content (e.g. when a picture has been photoshopped), and “virtual picture” or content for virtually created images (via AI for instance).

#### **4.6. Transparency obligations under the Proposal for an Artificial Intelligence Act**

In all the AI ethics charters and guidelines (Jobin et al. 2019) it is stressed the need to understand the decision-making processes of AI (Floridi et al. 2018) and the importance of a transparency principle (Pagallo-Durante 2022), articulated as the duty to make an object or entity knowable (Hayes (2020).

Following these recommendations, specific obligations to disclose AI influencers (see *supra* para. 2.1) shall enter into EU law when the AI Act is adopted, after the provisional agreement between the Parliament and the Council reached on 9 December 2023<sup>14</sup>.

The new rules establish the general principle according to which users should be made aware when they are interacting with AI, including systems that generate or manipulate image, audio, or video content, for example, deepfakes. In particular, Article 52(1) establishes that “providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system”<sup>15</sup>. The same is valid under para 3 for “an AI system that generates or manipulates image, audio, or video content that appreciably resembles existing persons...and would falsely appear to a person to be authentic or truthful (deep fake)”. As clarified by Recital 70, the purpose of such obligation is to “*take account of the specific risks of manipulation*”.

The user is requested to “disclose that the content has been artificially generated or manipulated”, but this is not an absolute requirement. Natural persons should be notified that they are interacting with an AI system “unless this is obvious from the circumstances and the context of use”.

This approach is however problematic because it leaves any evaluation to the user and adds a layer of subjectivity and uncertainty. Then the exception does not seem consistent with the rationale followed by the same EU legislator when establishing the duty to disclose the advertising nature of ad-contents (see *supra* para. 4.4). In that case, the notice that a content has been sponsored must be given even if it could be inferred by the context, for example because the sponsor company name is part of the message. The European Committee of the Regions pointed out this aspect in its Opinion over the AI

---

<sup>12</sup> Regulation (EU) 2022/2065 on a Single Market For Digital Services.

<sup>13</sup> Providers of online platforms are also requested to “provide recipients of the service with a functionality to declare whether the content they provide is or contains commercial communications”.

<sup>14</sup> Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending certain union legislative acts, COM/2021/206 final.

<sup>15</sup> This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate, and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.

Act proposal<sup>16</sup>, inviting to remove the exception on the grounds that “natural persons should always be duly informed whenever they encounter AI systems and this should not be subject to interpretation of a given situation. Their rights should be guaranteed at all times in interactions with AI systems”. Even if the final wording of the article has been modified after the amendments approved by the Parliament<sup>17</sup>, the exception is likely to be maintained.

At least, the definitive text is expected to better clarify what “notice” actually means. According to the revised text of para. 3 “Disclosure shall mean labelling the content in a way that informs that the content is inauthentic and that is clearly visible for the recipient of that content. To label the content, users shall take into account the generally acknowledged state of the art and relevant harmonised standards and specifications”. These references shall probably consent to adopt disclaimers in the form of hashtags followed by the notice, as widely used for disclosing the commercial nature of ad-contents, as well as to apply existing advertising self-regulations code of conduct, particularly the European Advertising Standard Authority (EASA) Best Practice Recommendation on Influencer Marketing and the similar provisions contained in local codes and national guidelines.

This obligation does not wave the EU institutions to set a similar obligation also for virtual influencers which does not technically amount to an AI system, but still may perfectly resemble a human and pose specific risks of deception.

## 5. LANDMARKS FOR AN ETHICAL ASSESSMENT

The assumption for establishing transparency obligations for virtual influencers is that virtual reality is fictional or illusory so what goes into it is not truly real. Many philosophers embrace this so called virtual fictionalism (Bateman 2011), according to which things that are supposed to happen in virtual words do not happen in reality.

The opposite virtual realism theory (Chalmers 2017) claims that virtual experiences are as valuable as non-virtual practices and are non-illusory. On this basis, the virtual realism theory is irreconcilable with any kind of discriminatory obligation imposed only on virtual influencers and not also on real-life ones.

This does not mean that virtual reality should be confined to something we can only perceive and describe. It is primarily something we can interact with and based on this, artificial agents can produce real consequences, including negative ones, to the human agents with which they interact (Durante 2020).

This work provides a concrete example and application of this theory, showing how the interaction between virtual influencers and consumers in digital environments, if does not transparently take place, is morally objectionable (Carson et al. 1985). Habermas’ critical social theory and Kantian ethics provide marketing communication researchers with a valuable theoretical perspective to impose information duties on such practices.

The basic idea of Habermas’ pragmatic theory of meaning is that “we understand a speech-act when we know what makes it acceptable” (Habermas 1984): to understand what a speaker means the hearer has to have access to the reasons for the speaker’s utterance. The idea is that agents, through their

---

<sup>16</sup> Opinion of the European Committee of the Regions — European approach to artificial intelligence — Artificial Intelligence Act (2022/C 97/12).

<sup>17</sup> The final draft should be written as follow “Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that the AI system, the provider itself or the user informs the natural person exposed to an AI system that they are interacting with an AI system in a timely, clear and intelligible manner, unless this is obvious from the circumstances and the context of use”.

speeches, make validity claims, i.e. a claim to truth, to rightness, and to truthfulness, sincerity, or authenticity, on which the meaning of their statements depends. These claims are inherent in all communicative actions. Therefore, the meaning of an utterance should not be found in the sentence itself, as per the encoding/decoding paradigm of language, but in the intent of the speaker and the reconstruction by the receiver, as per the intentionalist and dialogic paradigms (Kraus 1998). By loading language with intent, the speaker exercises an illocutionary force that can effectively change the hearer’s mind (Austin 1962) and eventually attain cooperation based on a shared understanding of reality. This process of interaction between subjects capable of speech and action that establishes interpersonal relations is that which Habermas calls communicative action. Such action can be triggered by potentially many different types of illocutionary forces, but especially by virtue of “shared knowledge, mutual trust, and accord with one another”.

Virtual marketing challenges this view because the utterances of virtual influencers do not imply validity claims, nor are trustworthy. The truthfulness and authenticity of recommendations shared by these influencers are highly questionable (Lou et al. 2023). In the end, virtual influencers can neither have actual experiences with products and services nor form their own opinions (Conti et al. 2022). Therefore, their endorsement of the product or service is in no way based upon its *bona fide* use, nor it is based upon personal opinions, beliefs, or experiences with or about the product or service because the virtual influencer has never used them.

As a result, without duties of disclosure, users are not enabled to detect and analyze distorted communications. This lack of disclosure impacts the “public sphere” (Habermas 1962; 1992), namely, the public arena wherein every democratic discourse is institutionalised. Since such a role, today is arguably played in large part by the Internet and social media, it seems fair to admit that placing boundaries to misleading communications plays a crucial role in the healthy functioning of a democracy.

Kantian ethics leads to a very similar conclusion. Within a Kantian system, businesses should not be able to use promotions that are dishonest or hide key facts (Perni 2023). The assumption hinges on Kant’s hypothetical publicity test in the second appendix to his Perpetual Peace. Even if specifically applied to the political sphere, the formula “all actions that affect the rights of other human beings, the maxims of which are incompatible with publicity, are unjust” (Kant 1795) has a normative and transcendental structure valid in any context concerning people interactions and coexistence (Pirni 2022). On this basis, we can critically assess current regulations and determine whether and to what extent they should be amended in accordance with duties of disclosure, publicity, and transparency.

## 6. CONCLUSIONS

Based on this ethical evaluation and review of the current legislation, I suggest that brands need to be transparent about using virtual characters in their communications through disclaimers. I advise that when using virtual influencers in advertising, brands should disclose this information because material to the purposes of the UCPD.

The question now is whether such disclosure is enough. My opinion is that whereby commercial communications contain a testimonial or endorsement of a product or service by the virtual influencer, a mere disclosure of virtuality does not prevent consumers’ deception. Indeed, in this case, the communication would not be, by definition, genuine, verifiable, and relevant, for the same reasons as those for which the communication lacks truth and authenticity under the Habermas’ validity claim test.

It is relevant in this connection that the EU Omnibus Directive prohibits fake reviews and endorsements (such as ‘likes’ on social media) of products and requires platforms to verify their authenticity and take reasonable and proportionate steps to ensure that these reviews are genuine and reflect the experience of real consumers. The Directive also establishes that traders giving access to such reviews should clearly state how the reviews are obtained and checked, and how they ensure that these come from consumers who have used or purchased the product (Durovic 2022).

The same approach guided advertising self-regulation authorities<sup>18</sup> to require that marketers must hold documentary evidence that a testimonial or endorsement used in a marketing communication is genuine, i.e. that the quote is from a real person, and it reflects what this said.

On these grounds, testimonials or endorsements by virtual influencers should be banned because of their lack of authenticity, while, in any other cases, the law should require a mandatory disclosure of their virtual nature. This kind of obligation is already placed by the AI Act Proposal for virtual influencers which do technically amount to an AI system and should be complemented by a similar duty of disclosure for those who are not AI-generated. These may still perfectly resemble a human and pose similar risks of impersonation or deception.

## REFERENCES

- Arsenyan, J., & Mirowska, A. (2021). Almost human? A comparative case study on the social media presence of virtual influencers. *International Journal of Human-Computer Studies*, 155(102694).
- Audrezet, A., de Kerviler, G., & Moulard, J. G. (2020). Authenticity under threat: When social media influencers need to go beyond self-presentation. *Journal of Business Research*, 117, 557-569.
- Austin, J.L. (1962). *How to Do Things with Words: The William James Lectures Delivered at Harvard University in 1955*, Harvard University Press.
- BEUC (2023). From Influence to responsibility. Time to regulate influencer marketing, BEUC-X-2023-093.
- BEUC (2022), “Dark Patterns” and the eu consumer law acquis. Recommendations for better enforcement and reform.
- Bateman, C. (2011). *Imaginary Worlds*. Zero Books.
- Boerman, S. C. (2020). The effects of the standardized Instagram disclosure for micro- and meso- influencers. *Computers in Human Behavior*, 103, 199-207.
- Callahan, K.. (2021). Cgi social media influencers: are they above the ftc's influence?. *Journal of Business and Technology Law*, 16(2), 361-386.
- Carson, T.L.; Wokutch, R.E.; Cox, J.E (1985), An Ethical Analysis of Deception in Advertising, *Journal of Business Ethics*, 1985, 4(2), 93-104.
- Cauffman, C.; Goanta, C. (2021). A new Order: The Digital Services Act and Consumer Protection. *European Journal of Risk Regulation*, 12(4), 758-774.
- Chalmers D.J. (2017). The Virtual and the Real. *Disputatio* 9(46), 309-352.
- Ciani, J., Tavella, M. (2017) La riconoscibilità della natura pubblicitaria della comunicazione alla prova del digital: native advertising tra obbligo di disclosure e difficoltà di controllo. *Informatica e diritto*, 2(1), 485.
- Conti, M., Gathani, J., & Tricomi, P. P. (2022). Virtual Influencers in Online Social Media. *IEEE Communications Magazine*, 60(8), 86-91.

---

<sup>18</sup> As the ICC under Article 13 Advertising and Marketing Communications Code or the UK Advertising Standard Authority – ASA under rule 3.45 of the UK Code of Non-broadcast Advertising and Direct & Promotional Marketing - CAP Code.

- Deng, F., & Jiang, X. (2023). Effects of human versus virtual human influencers on the appearance anxiety of social media users. *Journal of Retailing and Consumer Services*, 71, 103233.
- Dove, G., Halskov, K., Forlizzi, J., Zimmerman, J. (2017). UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17. New York, NY: Association for Computing Machinery. pp. 278–288.
- Durante, M. (2011). The Online Construction of Personal Identity through Trust and Privacy. *Information*, 2, 594-620.
- Durante, M. (2020). Technology and the Ontology of the Virtual. In S. Vallor (ed.) *The Oxford Handbook of Philosophy of Technology* (pp. 318–340). New York, NY: OUP.
- Durovic, M., Kniepkamp, T. (2022). Good advice is expensive – bad advice even more: the regulation of online reviews. *Law, innovation and technology*, 14(1), 128-156.
- EU Commission (2021). Guidance on the interpretation and application of Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market, C/2021/9320.
- EU Commission (2022) Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation. Final Report
- Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, 7(2), 117- 140.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E. (2018). AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, *Minds and Machines*, 28, 689-707.
- Franke, C., Groeppel-Klein, A., Müller, K. (2022). Consumers' Responses to Virtual Influencers as Advertising Endorsers: Novel and Effective or Uncanny and Deceiving? *J. of Advertising*, 1-17.
- Gill, R. (2023). Perfect. Feeling Judged on Social Media. New York, NY: Polity.
- Gunawan, D.D., Huarng, K.-H. (2015). Viral effects of social network and media on consumers' purchase intention. *Journal of Business Research*, 68, 2237-2241.
- Guthrie, S. (2020). Virtual influencers: More human than humans. In S. Yesiloglu & J. Costello (Eds.), *Influencer Marketing. Building Brand Communities and Engagement* (pp. 271- 285). Routledge.
- Habermas, J. (1992). Further reflections on the public sphere. *Habermas Public Sphere*, 428.
- Habermas, J. (1962). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, MIT Press.
- Habermas, J. (1984), *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*, Beacon Press, Boston, MA, 297.
- Hayes, P. (2020). An ethical intuitionist account of transparency of algorithms and its gradations. *Business Research*, 13, 849–874
- Hudson, S., Matson-Barkat, S., Pallamin, N., & Jegou, G. (2019). With or without you? Interaction and immersion in a virtual reality experience. *Journal of Business Research*, 100, 459-468.
- Hugh, D. C., Dolan, R., Harrigan, P., & Gray, H. (2022). Influencer marketing effectiveness: The mechanisms that matter. *European Journal of Marketing*, 56(12), 3485-3515.
- Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines, *Nature Machine Intelligence*, 1, 389–399.
- Kant, I. (1795), *Zum ewigen Frieden*, in *Kants Werke*, Berlin: Prussische Akademie Ausgabe, vol. VIII, 1923, 381.
- Krauss, R.M. & Chiu, C.-Y. (1998). Language and Social Behavior. In *Handbook of Social Psychology* 4<sup>th</sup> edn., Vol. 2, McGraw-Hill.

- Lou, C., Kiew, S.T.J., Chen, T., Lee, T.Y.M., Ong, J. E. C., & Phua, Z. (2022). Authentically Fake? How Consumers Respond to the Influence of Virtual Influencers. *Journal of Advertising*, ahead-of-print, 1-18.
- Luzak, J. (2015) Online Disclosure Rules of the Consumer Rights Directive: Protecting Passive or Active Consumers?. *J. Eur. Consumer & Mkt. L.*, 4(3), 79.
- Masteralexis, J., McKelvey, S., Keevan, S. (2021). #IAMAROBOT: Is It Time for the Federal Trade Commission to Rethink Its Approach to Virtual Influencers in Sports, Entertainment, and the Broader Market?, *Harvard Journal of Sports & Entertainment Law*, 2021, 12, 353-386, 376-377.
- Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, 56(4), 535–556.
- Moon, Y. (2003). Don't blame the computer: When self-disclosure moderates the self-serving bias. *Journal of Consumer Psychology*, 13(1-2), 125-137.
- Moustakas, E., Lamba, N., Mahmoud, D., & Ranganathan, C. (2020). Blurring lines between fiction and reality: Perspectives of experts on marketing effectiveness of virtual influencers. *International Conference on Cyber Security and Protection of Digital Services, Cyber Security 2020*, 1-6.
- Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A., & Dwivedi, Y. K. (2023). Deepfakes: Deceptions, mitigations, and opportunities. *Journal of Business Research*, 154.
- Mouritzen, S. L. T., Penttinen, V., & Pedersen, S. (2023). Virtual influencer marketing: the good, the bad and the unreal. *European Journal of Marketing*, Vol. ahead-of-print No. ahead-of-print.
- OECD (2022). Paper on Dark Commercial Patterns. Digital Economy Papers No. 336.
- Ohanian, R. (1991). The impact of celebrity spokespersons' perceived image on consumers' intention to purchase. *Journal of Advertising Research*, 31(1), 46-54.
- Pagallo, U., Durante M. (2022). The Good, the Bad, and the Invisible with Its Opportunity Costs: Introduction to the 'J' Special Issue on "the Impact of Artificial Intelligence on Law", *J*, 5(1), 139-149.
- Patnaik, P. (2021) Regulations for Social Media Influencers and Celebrity Endorsement, *Indian Journal of Law and Legal Research*, 3(1), 1-13.
- Penttinen, V., Ciuchita, R., & Čaić, M. (2022). YouTube It Before You Buy It: The Role of Parasocial Interaction in Consumer-to-Consumer Video Reviews. *Journal of Interactive Marketing*, 57(4), 561-582.
- Perni, R. (2023). Pubblicità, educazione e diritto in Kant, Firenze University Press.
- Pirni, A. (2022). At the roots of transparency: a public-ethics perspective. *Etica Pubblica*, 2, 15-27.
- Savageaux, L. (2022, August 18). Virtual Influencers: Harmless Advertising or Dystopian Deception?, *De Pauw The Prindle Institute for Ethics*.
- Woodroof, P. J., Howie, K. M., Syrdal, H. A., & VanMeter, R. (2020). What's done in the dark will be brought to the light: Effects of influencer transparency on product efficacy and purchase intentions. *Journal of Product & Brand Management*, 29(5), 675-688.