Deepfakes Are Coming: Does Australia Come Prepared?

Federica Celli*

The phenomenon of 'deepfakes' has increased exponentially since its first appearance in 2017 when several actresses saw their faces featured in adult movies without having given consent. The potential dangers to democracy as well as to individuals' reputation are significant. The article discusses the adequacy and applicability of Australian laws to deepfakes. It analysises the technology behind deepfakes and 'cheap fakes', followed by an examination of the misuse of deepfakes. It then offers a comparative analysis of European personality rights, the US right of publicity, and the US Deep Fakes Accountability Act before evaluating Australian tort, consumer and intellectual property law. It concludes that Europe's personality rights provides an effective and desirable legal response to deepfaking.

I INTRODUCTION

In his painting *ceci n'est pas une pipe*, Rene Magritte portrayed a pipe followed by the sentence 'this is not a pipe'. He demonstrated in a rather unique way that what human eyes see is not necessarily the truth. Although unique, anyone who is admiring Magritte's painting is ready to accept that what they see is indeed not a pipe, rather a painting depicting a pipe. Such an acceptance is not that automatic when looking at pictures or videos, which appear to resemble reality perfectly, yet they can be deceitful and portray something different from what is real. Technology is making that deception even easier, an example of that being the so-called 'deepfakes', which have been described as 'AI-generated media that depict made-up events ... [with] no agreed-upon technical definition'.[1]

Deepfakes can severely harm people's reputation and erode democracy, considering that '[f]alsehood diffuse[s] significantly farther, faster, deeper, and more broadly than the truth in all categories of information', particularly when it comes to false political news.[2] Since law tends to march with technology 'in the rear and limping a little,'[3] this article aims at assessing if and how badly Australian laws are currently limping in a potential march against deepfakes.

Such an evaluation appears crucial considering that even Facebook has launched a challenge to find the top-performing detection model for deepfakes,[4] which 'have proven to be something of an exaggerated menace for social media',[5] raising questions regarding what role such platforms should play in taking down deepfakes, which will not be the focus of this article.

The article firstly discusses deepfakes' origin and applications in Part II. Part III then examines the notion of personality rights in Europe and America as a possible response

[*] Federica Celli, MLS student at the University of Canberra.

[1] Joe Bateman, 'Deepfakes and Synthetic Media in the Financial System: Assessing threat Scenarios' (Working Paper No 7, Carnegie Endowment for International Peace, 8 July 2020) 4.

[2] Soroush Vosoughi, Deb Roy and Sinan Aral, 'The Spread of True and False News Online' (2018) 359(6380) *Science* 1146, 1146.

[3] *Mount Isa Mines Limited v Pusey* (1970) 125 CLR 383, 395 (Windeyer J).

[4] Christian Canton Ferrer et al, 'Deepfake Detection Challenge Results: An Open Initiative to Advance AI', *Facebook AI* (Web Page, 12 June 2020) <https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai>.

[5] James Vincent, 'Facebook Contest Reveals Deepfake Detection is Still an "Unresolved Problem"', *The Verge* (online at 12 June 2020) <https://www.theverge.com/21289164/facebook-deepfake-detection-challenge-unsolved-problem-ai>.

to deepfakes, as well as recent US deepfakes laws. Finally, Part IV discusses the application of current Australian laws as a response to deepfakes by also evaluating their adequacy and limits. Conclusively, Part V draws the article's final considerations.

II Deepfakes: How Did They Start And Where Are They Going

Deepfakes have their root in academic literature with a paper written in 2016. The authors' purported goal was to transfer the facial expressions of a source actor to a target actor by using a YouTube video portraying a facial performance of the latter in a way that 'it [was] virtually impossible to notice the manipulations.'[6] One year later, a Reddit user called 'deepfakes' posted several manipulated porn videos featuring celebrities by employing the open-source machine learning tool TensorFlow, and by compiling the celebrities' faces through 'Google image search, stock photos, and YouTube videos',[7] thus roughly applying what had been academically discussed the previous year. Such audio-visual manipulations became colloquially referred to as deepfakes.

Since emerging in 2017, deepfakes have rapidly developed in terms of numbers and perceived authenticity. In July 2019, with a 100% increase from 2018, the total number of deepfakes reached 14,678. 94% were pornographic videos with 134,364,438 views across only four dedicated deepfake pornography websites. [8] As of June 2020, deepfakes have had a 330% increase, reaching a total of 49,081.[9]

The next two paragraphs briefly address the differentiation between deepfakes and the so-called cheap fakes, followed by the possible uses and threats of the technology behind deepfakes.

A *Deepfakes and Cheap fakes*

Since 2017, the technology behind deepfakes has spread rapidly. Software such as Faceswap and Zao have allowed everyday-users to potentially create as well as distribute content comparable to more advanced deepfakes, resulting in what is called a democratisation of deepfakes' technology.[10] A differentiation between deepfakes and what is known as cheap fakes appears now relevant before continuing.

Cheap fakes arguably pre-date the digital age, an example being the fake bellicose telephone conversation between Margaret Thatcher and Ronald Reagan created by the anarcho-punk band Crass in the occasion of the 1983's UK elections.[11] A more recent example of a cheap fake is the video of Nancy Pelosi, the US House of Representatives Speaker, the speed of which was reduced in a way to make Mrs Pelosi sound drunk.[12]

[6] Justus Thies et al, 'Demo of Face2Face: Real-time Face Capture and Reenactment of RGB Videos' (Conference Paper, ACM SIGGRAPH 2016 Emerging Technologies, July 2016) 2387.

[7] Samantha Cole, 'AI-assisted fake porn is here and we're all fucked', *Vice* (online at 12 December 2017) <https://www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn>.

[8] Deeptrace, *The State of Deepfakes: Landscape, Threats, and Impact* (Report, September 2019) 1 <https://sensity.ai/reports/#> ('*The State of Deepfakes*').

[9] Henry Ajder, 'Deepfake Threat Intelligence: a Statistics Snapshot from June 2020', *Sensity Formerly Deeptrace* (Web Page, 3 July 2020) <https://sensity.ai/deepfake-threat-intelligence-a-statistics-snapshot-from-june-2020/>.

[10] Hannah Smith and Katherine Mansted, *Weaponised Deep Fakes: National Security and Democracy* (Report No 28/2020, April 2020) 5.

[11] Ibid 6; TheRobman, 'Crass Thatchergate Tape' (YouTube, 24 June 2012) <https://www.youtube.com/watch?v=QmfLP1IOip8>.

[12] Washington Post, 'Pelosi Videos Manipulated to Make Her Appear Drunk Are Being Shared on Social Medial' (YouTube, 24 May 2019) <https://www.youtube.com/watch?v=sDOo5nDJwgA&t=46s>.

The cheap fake-deepfakes spectrum helps distinguish those two phenomena. Their differences primarily concern technical sophistication, techniques, and barriers to entry.[13] Where deepfakes represent the end of the spectrum characterised by a higher computationally reliance and the least publicly accessibility,[14] cheap fakes refer to techniques such as photoshopping, slowing and speeding moving images, recontextualization, and lookalikes.[15] More generally, cheap fakes, sometimes also known as 'shallowfakes' represent those manipulations that do not rely on AI.[16]

Although this article specifically addresses deepfakes, most of the following considerations arguably apply to cheap fakes as well. Indeed, if on the one hand Generative Adversarial Networks (GANs), which is a machine learning ('ML') configuration where two MLs systems compete to improve the learning of a task, are contributing to the increased ability to generate even more convincing deepfakes,[17] on the other hand cheap fakes also are increasingly becoming more credible with the advancement of technology.

Finally, it is worth mentioning that social media apps are increasingly using AI technologies to engage their users by, for instance, allowing them to manipulate videos with a smartphone. An example of such an app is Reface, which is currently ranked number eight in the top charts for free entertainment apps.[18] Setting aside the possible issues in relation to privacy, Reface allows its users to put their faces on popular GIFs by simply using one of their pictures, and it also offers a premium option which allows users to 'upload their own, choose faces from their gallery', and providing results without watermarks or adverts.[19] Although the final result is arguably still easily recognisable as having been manipulated, the quality of the final GIF can be quite convincing. Furthermore, the ability to simply use one photo to create the manipulated video makes it extremely easy to create a GIF without the consent of the person involved, considering that Facebook is a great and accessible resource when it comes to people's pictures.

B *Deepfakes' Applications and the 'Liars Dividend'*

It has been said that for human beings to survive and improve, they need to constantly acquire knowledge about the world by relying on trustworthy sources of information such as direct visual perception.[20] However, because it is impossible to always witness in first person an event occurring, videos represent 'the next best thing'.[21] The same principle applies to hearing, which is another sense that generally leads to trust and believe in what is heard whether that would be in person, over the phone or, again, through a video. Deepfakes are arguably demolishing the trustworthiness of both sight and hearing.

[13] Britt Paris and Joan Donovan, Data & Society, *Deepfakes and Cheap fakes: The Manipulation of Audio and Visual Evidence* (Report, 18 September 2019) 10-11.
[14] Ibid.
[15] Ibid 24.
[16] Bateman (n 1) 28.
[17] M Caldwell, J T A Andrews, T Tanay and L D Griffin, 'AI-enabled Future Crime' (2020) *Crime Science* 1, 5.
[18] 'Top Charts Ranking for the App Store for iPhone apps', *Appfollow* (Web Page, 25 August 2020) <https://appfollow.io/rankings/iphone/au/entertainment#2020-10-21> (last accessed 27 October 2020).
[19] Adam Smith, 'Reface: New App Lets You Put Your Face on GIFs – But Is It Safe?', *Independent* (online at 7 August 2020) <https://www.independent.co.uk/life-style/gadgets-and-tech/news/doublicat-gif-safe-app-deep-fake-celebrities-a9628906.html>.
[20] Don Fallis, 'The Epistemic Threat of Deepfakes' [2020] (August) *Philosophy & Technology* 1, 2 <https://link.springer.com/article/10.1007/s13347-020-00419-2>.
[21] Ibid.

The technology behind deepfakes has also been employed for noble purposes. For instance, the Scottish company CereProc offers, among other services, Cerewave AI, a neural text-to-speech system powered by the same deepfakes' machine learning technology, which has also been used to produce a reconstruction of the original voice of a customer, who suffered from a degenerative illness, by fusing the customer's voice with a relative's voice.[22] However, the same technology was employed to fraud and steal money. Indeed, an energy company's CEO transferred a considerable amount of euros to a Hungarian supplier, believing that was at the direction of his parent company's chief executive, who, unfortunately, had never given that order, the CEO being the victim of a 'voice-spoofing attack' made possible by AI.[23] This 'deepfake vishing (voice phishing)' is easily employable in identity theft, requiring a small amount of audio data to be created, making it possible to clone someone's voice by using a social media clip or voicemail greetings.[24]

David Beckham's manipulated videos are another example of benign employment of deepfakes' technology, which spread awareness about the Malaria Must Die initiative by manipulating Beckham's original video where he spoke English, to allow him to speak nine different languages without needing Beckham actually to learn those languages.[25]

In 2018, some researchers at Berkeley had used AI to transfer professional dance moves from a professional dancer to a target person.[26] This innocuous deepfake arguably discloses how the step towards harmful employment of the technology is quick to make. Indeed, as Agnieszka Walorska highlighted, '[w]hat would, for instance, be the ramifications of a video showing a politician performing a Nazi salute or even just giving the middle finger?'.[27] The answer remands to deepfakes' ability to compromise democracy and spread misinformation.

Such a threat is perceived now more than ever, when, in the middle of a pandemic, the internet has had an increase in demand, which was up to 80% in April 2020 to the point it caused 'the worst average internet congestion' in some cities.[28] With more people spending time over the internet to get entertained as well as informed, the possibility to encounter deepfake videos, and potentially believe in the truthfulness of those videos at least initially, increases exponentially.

With the upcoming US election, the threat deepfakes poses to democracy has prompted corporations such as Sensity, 'the world's first visual threat intelligence company',[29] to provide to its subscribers a 'special fake video monitor for US2020' giving daily updates on the incidence of deepfakes detected by Sensity's Platform, targeting US candidates

[22] 'Cerewave AI,' *CereProc* (Web Page) <https://www.cereproc.com/en/v6>.

[23] Catherine Stupp, 'Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case' *Wall Street Journal* (online at 30 August 2019) <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>.

[24] Bateman (n 1) 9-11.

[25] Mike Butcher, 'The Startup Behind that Deep-fake David Beckham Video Just Raised $3' *TechCrunch* (Web Page, 25 April 2019) <https://techcrunch.com/2019/04/25/the-startup-behind-that-deep-fake-david-beckhamvideo-just-raised-3m/> accessed 26 November 2019.

[26] Caroline Chan et al, 'Everybody Dance Now' (Conference Paper, IEEE/CVF International Conference on Computer Vision, 7 August 2019) 5932-5941 <https://arxiv.org/pdf/1808.07371.pdf>.

[27] Agnieszka M Walorska, *Deepfakes and Disinformation* (Friedrich Naumann Foundation for Freedom, May 2020) 16.

[28] Nick Sas, 'Coronavirus affecting internet speeds, as COVID-19 puts pressure on the network', *ABC News* (online at 1 April 2020) <https://www.abc.net.au/news/2020-04-01/coronavirus-internet-speeds-covid19-affects-data-downloads/12107334>.

[29] 'Defending Individuals and Organizations from Visual Threats', *Sensity Formerly Deeptrace* (Web Page) <https://sensity.ai/about/>.

to the Presidential elections as well as other people of interest. As of 19 October 2020, Donald Trump is the most targeted candidate with 174 deepfakes and four cheap fakes.[30] Although those videos do not appear to truly compromise the political balances unless Trump announcing 'the Milky Act' is considered a severe threat to democracy,[31] examples from the past show how deepfakes can create political turbulences. For instance, in June 2019 a sex tape portraying a Malaysian minister engaging in homosexual sexual intercourse compromised the minister's political career and risked him facing criminal charges, homosexuality being illegal in Malaysia.[32]

The Malaysian minister's video has not officially been declared to be a deepfake although the minister claimed so.[33] This situation shows the so-called liar's dividend problem arising from deepfakes. Since the quality of deepfake's videos is continuously getting better, it is possible that people caught doing something wrong might invoke the deepfake card to get away with that misbehaviour. Moreover, as concisely stated by Professor Citron '[r]egrettably and perversely, the Liar's Dividend grows in strength as people learn more about the dangers of deep fakes'.[34]

To raise awareness, Sensity has created with Microsoft and the University of Washington's Centre for an Informed Public the project 'Spot the Deepfake', a quiz which goal is to educate people about deepfakes,[35] to have 'informed citizens who critically question what they see and hear, and who are looking for confirmation that the media has been vetted'.[36]

While the threat to democracy represents a real risk, as of June 2020, only 4% of deepfakes targeted people with political backgrounds,[37] pornography accounting for 96% of deepfakes,[38] including pornographic deepfakes featuring celebrities as well as revenge porn content. Revenge porn resulting from image manipulation already existed as 'sexualised photoshopping', i.e. the superimposition of a pornographic image onto a person's head or body.[39] Although the medium has been refined and turned into a proper video rather than a mere photoshopped image, the point is still that 'the fact that an image [or video] has been altered, or is even composed of images taken of different women, does not diminish the potential harm resulting from its dissemination'.[40]

[30] The data can be found at this link <https://platform.sensity.ai/us2020>, provided you have signed up for Sensity.

[31] Mista Jones, 'President Trump Announces the Milky Act' (YouTube, 5 October 2020) <https://www.youtube.com/watch?v=QPWyg8nbeoo>.

[32] Jarni Blakkarly, 'A Gay Sex Tape Is Threatening to End the Political Careers of Two Men in Malaysia', *SBS News* (online at 17 June 2019) <https://www.sbs.com.au/news/a-gay-sex-tape-is-threatening-to-end-the-political-careers-of-two-men-in-malaysia>.

[33] Ibid.

[34] *The National Security Challenge of Artificial Intelligence, Manipulated Media, and "Deep Fakes" Before the H. Permanent Select Comm. on Intelligence*, 116th Cong. 7 (2019) (statement of Danielle Keats Citron, Morton & Sophia Macht Professor of Law, University of Maryland Carey School of Law).

[35] Giorgio Patrini, 'How to Spot a Deepfake: Educating the Public on Deepfakes', *Sensity Formerly Deeptrace* (Web Page, 8 October 2020) <https://sensity.ai/how-to-spot-a-deepfake-educating-the-public-on-deepfakes/>.

[36] Ibid.

[37] Ajder (n 9).

[38] *The State of Deepfakes* (n 8) 1-2.

[39] Clare McGlynn, Erika Rackley, and Ruth Houghton, 'Beyond "Revenge Porn": The Continuum of Image-Based Sexual Abuse' (2017) 25(1) *Feminist Legal Studies* 25, 33.

[40] Australian Women Against Women Alliance, Submission No 19 to Senate Legal and Constitutional Affairs References Committee, Parliament of Australia, *Inquiry into the Phenomenon Colloquially Referred to as 'Revenge Porn'* (14 January 2016) 6.

What are legal strategies to combat the unauthorised manipulation of someone's image, primarily focusing on the civil rather than criminal aspects?

III Possible Legal Responses Applicable to Deepfakes

As previously stated, deepfakes' harm involves the manipulation of someone's likeness, image, and voice without that person's consent. Consequently, any adequate legal response to deepfakes must consider such a non-consensual element.

The following paragraphs examine the so-called personality rights, which Australia, following the UK, does not specifically recognise,[41] and some US acts and bills on deepfakes, with particular attention on the DEEP FAKES Accountability Act ('DF Act').[42]

A *Personality Rights: the European and United States Approaches Compared*

Personality rights are said to be 'a civil law concept',[43] seen as a bundle of rights that includes the right to one's name, image, right to privacy, and in general it relates to the protection of individuals' integrity and inviolability.[44] German scholars were among the first to use the word *Persönlichkeitsrecht* during the nineteenth and twentieth centuries.[45] Although the concept subsequently spread also to extra-European countries,[46] in Anglo-American countries the term 'personality' does not connotate the protection of people's bodily and non-bodily aspects.[47] The US has adopted the different concept of 'right to publicity', which is more connected to the idea of protecting liberty rather than protecting individuals' dignity as pursued by the European concept of personality rights.[48] Since the latter is said to have derived from the 'particular synergy' between German and French scholars and case law,[49] those countries are here used as representative of the European approach to personality rights.

As mentioned previously, deepfakes' harm comes from the unconscionable use of an individual's likeness perpetrated by a third person. French law protects from such harm by recognising that individuals have an exclusive right to their image and its use, which embodies the power to prohibit its dissemination and reproduction when that occurs without express and specific permission.[50] France had recognised the right of a person to their own likeness already in 1858 when the family of the actress Rachel was awarded damages as a result of the unauthorised publication of the actress' portraits over her deathbed.[51] France's personality rights are broadly interpreted, they are reflected in the

[41] See generally Rosina Zapparoni, 'Propertising Identity: Understanding the United States Right of Publicity and Its Implications - Some Lessons for Australia' (2004) 28(3) *Melbourne University Law Review* 690.

[42] DEEP FAKES Accountability Act, HR 3230, 116th Congress (2019) ('DF Act').

[43] Giorgio Resta, 'Personnalite, Personlichkeit, Personality' (2014) 1(3) *European Journal of Comparative Law and Governance* 215, 216-217 [1].

[44] Ibid.

[45] Adrian Popovic, 'Personality Rights – a Civil Law Concept' (2004) 50(2) *Loyola Law Review* 349, 351.

[46] See, eg, *Código Civil* [Civil Code] (Brazil) Ch II Book 1 Title 1; *Civil Code of Quebec*, CQLR c CCQ-1991 Art 3.

[47] Resta (n 43) 222 [2].

[48] Robin D Barnes, *Outrageous Invasions: Celebrities' Private Lives, Media and the Law* (Oxford University Press, 2010) 15.

[49] Resta (n 43) 228 [5].

[50] Gert Bruggemeier, Aurelia Colombi Ciacchi and Patrick O'Callaghan (eds), *Personality Rights in European Tort Law* (Cambridge University Press, 2010) 284.

[51] *Rachel*, Tribunal Civil de la Seine (Paris), DP 1858.3.62, 16 June 1858.

words of article 9 of the *Code Civil*, which states that everyone has the right to see their private life respected, and empowers courts to issue injunctions to stop or prevent intrusions into the intimacy of one's private life.[52] The protection is independent of both the notoriety and profession of the person, always requiring a specific authorisation to use the person's photograph.[53]

Germany impliedly recognises a right of personality in its Constitution, Civil Code and its 1907 Act on the Protection of Works of Art and Photographs.[54] Germany's right of personality is conceived as a unitary protection of patrimonial and non-patrimonial interests, where the autonomy of self-determination represents the core.[55] A landmark case, which showed that 'the notion of human dignity is the conceptual and normative backbone of all German constitutional law',[56] upheld an injunction preventing the printing, distribution, and publishment of a novel which character was modelled on the author's brother-in-law, although named differently and more caricatured.[57] The injunction's basis was that freedom of art presupposes dignity, which represents the 'supreme and controlling value of the whole system of basic rights'.[58] Another interesting case, which perfectly shows the applicability of personality rights' protection to deepfakes, is the one where a German goalie successfully stopped FIFA videogame from using his likeness and name without his explicit consent,[59] the court interestingly stating with regard to the use of the goalie's name that the damage derived from the player's infringed right to choose how his name might be used.[60]

Differently from France and Germany, America has adopted the narrower concept of 'right of publicity'.[61] Born from the right 'to be left alone' conceived in 1890 to protect publications of individual's private matters when publicly irrelevant,[62] the term right of publicity was arguably firstly mentioned[63] when a baseball player successfully won a case against the use of his image on baseball cards, considered as merchandise exploiting his image for profit.[64] Such a right has been described as one 'inherent to everyone to control the commercial use of identity and persona',[65] so disclosing its primarily patrimonial focus. Moreover, differently from Germany and the protection

[52] *Code Civil* [Civil Code] (France). See also Elisabeth Logeais and Jean-Baptiste Schroeder, 'The French Right of Image: An Ambiguous Concept Protecting the Human Persona', [1998] 18 *Loyola of Los Angeles Entertainment Law Review* 511, 515.

[53] Bruggemeier, Colombi Ciacchi and O'Callaghan (n 50) 285.

[54] Carol J Greer, 'International Personality Rights and Holographic Portrayals' (2017) 27(2) *Indiana International and Comparative Law Review* 247, 265.

[55] Tatiana Synodinou, 'Image Right and Copyright Law in Europe: Divergences and Convergences' (2014) 3 *Laws* 181, 185.

[56] Hannes Rosler, 'Dignitarian Posthumous Personality Rights – An Analysis of U.S. and German Constitutional & Tort Law', (2008) 26(1) *Berkeley Journal of International Law* 153, 168.

[57] *Mephisto*, Bundesverfassungsgericht [German Constitutional Court], 1 BvR 435/68, 24 February 1971 reported in 30 BVerfGE 173.

[58] Ibid.

[59] *Oliver Kahn v Electronic Art*, Oberlandesgericht Hamburg [Hamburg Court of Appeal], 7 U 41/03, 13 January 2004.

[60] Greer (n 54) 268.

[61] Ibid 256.

[62] Samuel D Warren and Louis D Brandeis, 'The Right to Privacy', [1890] 4 *Harvard Law Review* 193, 195, 206.

[63] Melville B Nimmer, 'The Right of Publicity' [1954] (Spring) *Law and Contemporary Problems* 203, 204, 218-23. Cf Jennifer E Rothman, 'The Right of Publicity's Intellectual Property Turn' (2019) 42(3) *Columbia Journal of Law & the Arts* 277, 281-288 [I].

[64] *Haelan Laboratories v Topps Chewing Gum*, 202 F 2d 866 (2nd Cir, 1953). See also William K Ford and Raizel Liebler, 'Games Are Not Coffee Mugs: Games and the Right of Publicity' (2012) 29(1) Santa Clara High Technology Law Journal 1, 7-8.

[65] J Thomas McCarthy and Roger E Schechter, *The Rights of Publicity & Privacy, 2d* (Thomson Reuters, 2nd ed, 2020) vol 1, [1:3].

given by the broader personality right concept in Europe, the American First Amendment has mostly prevailed over the right of publicity's protection of people's likeness and name when claimed against videogames.[66] Indeed, videogames in the US have been considered to represent a form of expression similar to books.[67]

US publicity rights are stated differently in each state with California arguably representing the 'golden standard'[68] by recognising such rights both statutorily and at common law.[69] Moreover, differently from the general US trend mentioned above, California had also enforced the protection of publicity rights against videogames.[70] Nevertheless, even California's publicity rights 'almost always fail to overcome' the First Amendment's defence,[71] arguably showing that the broader concept of personality rights linked to individuals' dignity guarantees greater protection and would better tackle deepfakes.

B *The US Deepfakes-Targeted Laws*

Although several American states had passed deepfakes-targeted laws, those are still affected by the rights of free speech under the First Amendment.[72] California has, for instance, passed two bills in 2019 amending various sections of the Civil Code and Code of Civil Procedure to prevent both the improper influence of elections and unauthorised use of people's likeness in pornography that deepfakes could cause.[73] Similarly, Virginia has criminalised maliciously distributed unauthorised pornographic material, while Texas specifically criminalised deepfakes aimed at injuring political candidates or influencing elections' results, both states providing incarceration for the laws' violation.[74]

At the federal level, the DF Act[75] is currently before the Congress. Such Act requires anyone sharing a harmful deepfake to disclose it is fake in a way which varies depending on the medium.[76] It provides about the establishment of the Deep Fakes Task Force to detect deepfakes and differentiate them from legitimate audio-visual

[66] *CBC Distribution and Marketing v Major League Baseball Advanced Media*, 505 F 3d 818 (8th Cir, 2007).

[67] *Brown v Entertainment Merchants Association*, 564 US 786 (2011).

[68] Kelsey Farish, 'Do deepfakes pose a golden opportunity? Considering whether English law should adopt California's publicity right in the age of the deepfake' (2020) 15(1) *Journal of Intellectual Property Law & Practice* 40, 46.

[69] The right was firstly established in *Eastwood v Superior Court* (respondent National Enquirer, Inc., as real party in interest), 149 Cal App 3d 409 (Cal Ct App, 1983) at common law, and it is recognised at a statutory level by §3344 of Cal Civ Code (Deering 2020), which states '(a) [a]ny person who knowingly uses another's name, voice, signature, photograph, or likeness, in any manner, on or in products, merchandise, or goods, or for purposes of advertising or selling, or soliciting purchases of, products, merchandise, goods or services, without such person's prior consent, or, in the case of a minor, the prior consent of his parent or legal guardian, shall be liable for any damages sustained by the person or persons injured as a result thereof …'.

[70] *No Doubt v Activision Publishing, Inc*, 122 Cal Rptr 3d 397 (Cal Ct App, 2011).

[71] Farish (n 68) 46.

[72] Penelope Thornton et al, 'Deepfakes: an EU and US perspective' [2020] (July) *Lexology* 3 <https://www.lexology.com/library/detail.aspx?g=8f038b17-a124-46b2-85dc-374f3ccf9392>.

[73] KC Halm, Ambika Kumar Doran, Jonathan Segal and Caesar Kalinowski IV, 'Two New California Laws Tackle Deepfake Videos in Politics and Porn', *Davis Wright Tremaine LLP* (Web Page, 14 October 2019) <https://www.dwt.com/insights/2019/10/california-deepfakes-law>.

[74] Thornton et al (n 72) 3-4.

[75] Also before the congress since 2018 is the Malicious Deep Fake Prohibition Act of 2018, S 3805, 115th Congress (2018), which will not be discussed in this article.

[76] DF Act (n 42) s 2 §1041.

recordings,[77] and prescribes criminal liability on pornographic deepfakes' creators as well as an in rem civil action against the content when the author is unknown.[78] Although the bill appears overall thorough, it has been criticised for not truly discouraging malicious deepfakes' creators, and because the term 'deepfakes' is so overbroad that it would be more sensible to let courts apply current laws to address each particular case better rather than using a standardised approach.[79] Additionally, it has been highlighted how criminalisation might not be appropriate, considering the complexity and ubiquity of deepfakes' technology and their ability also to facilitate free speech, and social and political comment.[80]

IV AUSTRALIAN LEGISLATION

Australia does not specifically recognise personality rights, nor has it created anything similar to the American right of publicity. Thus, in addressing whether current Australian laws can adequately respond to the threat posed by deepfakes, different statutes and case laws must be analysed. Specifically, Australian consumer and intellectual property laws, as well as torts laws such as passing off and defamation, could represent valuable means to fight deepfakes.

The following paragraphs discuss the application of such laws to deepfakes, and whether they provide sufficient protection to individuals who see their images unauthorizedly manipulated.

A *Copyright, Trademarks, and Consumer Laws*

Especially with regard to celebrities, copyright, trademarks and consumer laws could successfully protect against deepfakes. According to the *Copyright Act*,[81] celebrities are generally considered the owner of their performance, hence arguably of the image and voice related to that performance, potentially allowing them to sue for any infringement of their copyright caused by a deepfake which unauthorizedly manipulated their image, and seek relief through injunctions, damages or account of profits.[82]

It should be briefly mentioned that also the author of a deepfake could, at least in theory, claim copyright authorship. Indeed, deepfakes could arguably be considered 'original artistic work',[83] for instance, and confer to their creators an exclusive right to reproduce, publish and communicate the work, i.e. the deepfake, to the public.[84] The situation could be complicated further when the image used to create the deepfake video is subjected to third-party copyright, for instance, the professional photographer who took the photo, who would then arguably be the one entitled to sue the deepfake creator. In this scenario, if the third-party has authorised to make the deepfake or made it themselves, it might be difficult for the person portrayed in the manipulated image to find protection.

Furthermore, since celebrities' images can be highly valuable, being able to influence consumer choices if used in products' advertisements, trademarks are being used as a

[77] Ibid s 7.

[78] Ibid ss 2, 4.

[79] Zachary Schapiro, 'DEEP FAKES Accountability Act: Overbroad and Ineffective' [2020] (April) *Boston College Intellectual Property & Technology Forum* 1, 15.

[80] See Tyrone Kirchengast, 'Deepfakes and Image Manipulation: Criminalisation and Control' (2020) 29(3) *Information & Communications Technology Law* 308.

[81] *Copyright Act 1968* (Cth) ('*Copyright Act*').

[82] Ann Slater, 'Personality Rights in Australia' (2001) 20(1) *Communications Law Bulletin* 12, 12.

[83] *Copyright Act* (n 81) s 32.

[84] Ibid s 31(b).

mean to control celebrities' personalities.[85] Under the *Trade Marks Act*, for a successful registration, the trademark must be capable of distinguishing the applicant's goods or services in respect of which the trademark is sought,[86] an example could be Paul Newman's sauces which have his face and signature on the label.[87] However, arguably a celebrity could trademark their image even when there is not such a connection with a good, changing the nature of trademark protection and assimilating it to a generalised image or personality protection.[88] Thus, since a trademark infringement occurs when there is a deceptive and almost identical use of the image,[89] deepfakes would easily represent such an infringement.

Additionally, a celebrity's image could also be protected through consumer laws, specifically when it is used in a misleading and deceptive way,[90] for instance the unauthorised display of a music group on a t-shirt[91] or the photograph of a professional swimmer used in a way as to falsely suggest he sponsored a telephone company.[92] Hence, a deepfake created to advertise a specific service or product without the actual authorisation of a certain celebrity could be stopped under consumer laws.

Nevertheless, copyright, trademarks and consumer laws, while arguably applicable to celebrities due to their notoriety, would hardly apply to ordinary people, since the damage created by the unauthorised use of the celebrity's image derives from the profit related to that same image because famous, which would not be the case for non-famous people.

B *Defamation and Passing off*

Defamation is a tort to protect individuals' reputation from being diminished in the eyes of the public due to someone else's false assertions,[93] regardless of whether that someone intended to have that effect and had, indeed, reasonable care.[94]

For an action in defamation to succeed, there must be three elements. Firstly, provided the victim is alive,[95] the defamatory matter must be published, it must be communicated to someone other than the plaintiff.[96] Deepfakes spread around the internet, on websites and social media, hence this element would easily be present, considering that Facebook posts have already been held capable of representing defamatory matters.[97] Secondly, the defamed must be identifiable, which is satisfied by merely portraying a person on television or in a photograph without needing to name them.[98] Deepfakes are all about images manipulation, thus the second element is connatural in such videos. Finally, the matter must be defamatory in the eyes of a right-

[85] Lynne Weathered, 'Trade Marking Celebrity Image: the Impact of Distinctiveness and Use as a Trade Mark' (2000) 12(2) *Bond Law Review* 161, 163.

[86] *Trade Marks Act 1995* (Cth) s 41(2) ('*Trade Marks Act*').

[87] Weathered (n 85) 167.

[88] Ibid 167-168.

[89] *Trade Marks Act* (n 86) s 120.

[90] *Competition and Consumer Act 2010* (Cth) sch 2 s 18.

[91] *Hutchence v South Seas Bubble Co Pty Ltd* (1986) 64 ALR 330.

[92] *Talmax Pty Ltd v Telstra Corporation Ltd* (1997) 2 Qd R 444. See generally Pauline Sadler, 'Character Merchandising and the Sporting Industry' [2001] 3 *Legal Issues in Business* 57.

[93] Peter Nygh and Peter Butt, *Butterworths Australian Legal Dictionary* (Butterworths, 1997) 333.

[94] *Dow Jones & Co Inc v Gutnick* (2002) 210 CLR 575 [25].

[95] *Civil Law (Wrongs) Act* 2002 (ACT) s 122.

[96] *Pullman v Walter Hill & Co* [1891] 1 QB 524, 537 (Lord Esher MR).

[97] *Reid v Dukic* [2016] ACTSC 344.

[98] *Johnston v Australian Broadcasting Commn* (1993) 113 FLR 307; *Nixon v Slater & Gordon* (2000) 175 ALR 15.

thinking member of the society.[99] This could at times not be present in deepfakes, defamation requiring some sort of actual damage to the individual's reputation, and not protecting the mere unauthorised use and distribution of that individual's image.

Moreover, even when damages have occurred, while Australian differently from the US does not have an equivalent of the freedom of speech, the common law defence of opinion could successfully be used when the deepfake is clearly satirical.[100]

Another possible protection could be the tort of passing off. Briefly, such tort arises when a person's reputation is misrepresented by another person so to create a damage to the first person's business.[101]

Courts had upheld passing off in cases of unauthorised use of the plaintiff's photograph[102] or created personality[103] when that gave the public the erroneous impression that the plaintiff consented to such use and profited from it. It was also upheld for the publication of a written work's parody when there was no sufficient indication that it was simply a satirical copy of the original.[104] Indeed, the tort has been also used for 'authors protection of their goodwill from damage occasioned by false attribution of authorship'.[105]

Thus, arguably such a tort could be used against deepfakes. For instance, it could be used in the case of a politician's video manipulated as to create a parody, when its satirical aspect is not that obvious to the public, and could fuel misinformation about that politician's views or his or her reputation. This could be made possible by loosely interpreting the word 'authors' as to include the protagonists of the original video, which was then manipulated.

Nevertheless, once again a commercial aspect must be present, courts having dismissed cases where, although a famous person's image was used without that person's consent, the use of that image did not induce the public to think there was a commercial connection between such person and the company or product which exploited that person's image.[106]

V CONCLUSION

Deepfakes represent a serious problem which is growing consistently. From the considerations above, the protection afforded by current laws appears primarily commercial-related, Australia evidently lacking in providing a protection to people's likeness, image and voice in a broader sense.

The analysis of Europe and the US shows that the most desirable approach would be to recognise a type of personality rights similar to the one conceived by France and Germany, the American right of publicity being, similarly to Australia, anchored in commercial considerations. The DF Act also does not represent the best solution, risking to compromise the 'bad' as well as the 'good' deepfakes. Indeed, as wisely

[99] *Reader's Digest Services Pty Ltd v Lamb* (1982) 150 CLR 500, 506; *Sim v Stretch* (1936) 52 TLR 669, 671.

[100] *Seidler v John Fairfax* (1986) Aust Torts R 80-002.

[101] R P Balking and J L R Davis, *Law of Torts* (LexisNexis Butterworths, 5th ed, 2013) 719 [23.7].

[102] *Henderson v Radio Corporation Pty Ltd* (1960) 60 SR (NSW) 576.

[103] *Hogan v Koala Dundee Pty Ltd* (1988) 83 ALR 187.

[104] *Clark v Associated Newspapers Ltd* [1998] 1 All ER 959.

[105] Ibid [18].

[106] See, eg, *Newtown-John v Scholl-Plough (Australia) Ltd* (1986) 11 FCR 233; *Honey v Australian Airlines* (1990) 18 IPR 185.

pointed out by Lawrence Lessig, '[law] should regulate culture only where that regulation does good'.[107]

To conclude, Australian laws are limping in their march against deepfakes. However, personality rights, if specifically recognised, could be Australia's clutches, at least temporarily.

***

[107] Lawrence Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* (Penguin Press, 2004) 305.