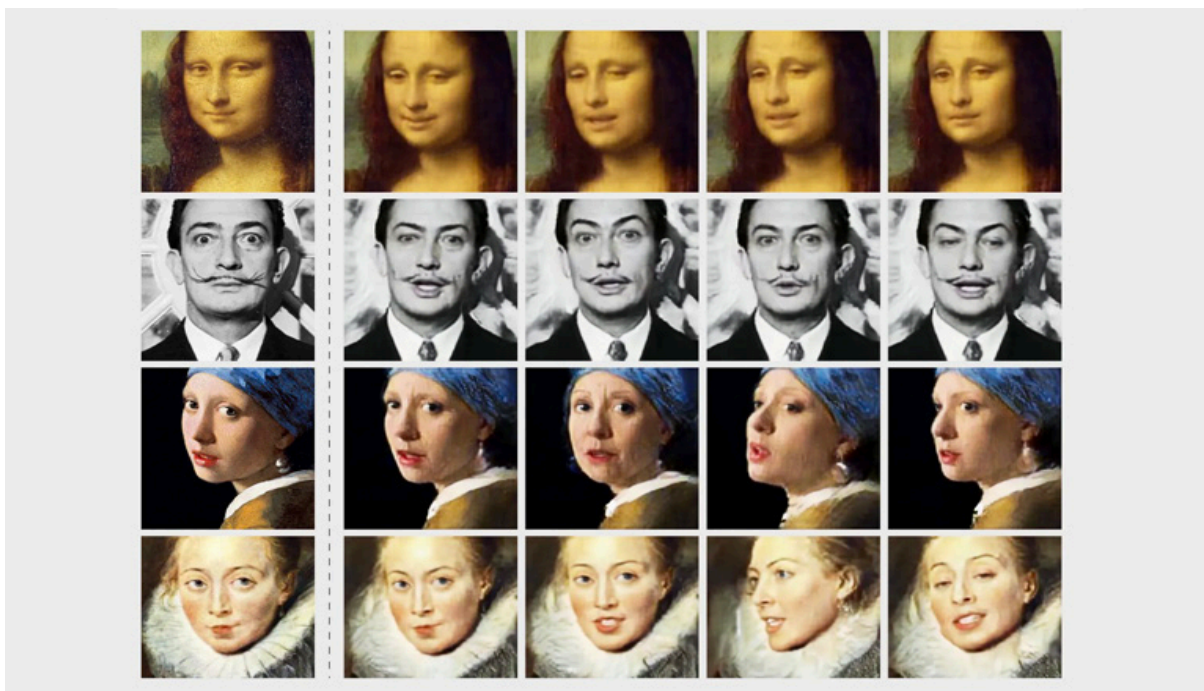


# Η τεχνολογία των deepfakes και οι επιπτώσεις τους στη ζωή μας

Ειρήνη Βασιλοπούλου

Δικηγόρος, Υπ. ΔΝ ΕΚΠΑ



**Η** συνεχής πρόοδος της επιστήμης και της τεχνολογίας, ειδικότερα με την ανάπτυξη και χρήση της τεχνητής νοημοσύνης (artificial intelligence), οδήγησαν πριν από λίγα χρόνια και στη δημιουργία των λεγόμενων deepfakes, τα οποία αναμφισβήτητα αποτελούν παγκοσμίως μια σύγχρονη πρόκληση.

Τι είναι όμως η τεχνολογία των deepfakes; Τα deepfakes είναι βίντεο στα οποία ένα πρόσωπο εμφανίζεται να λέει ή να κάνει πράγματα τα οποία ουδέποτε είπε ή έκανε. Δηλαδή πρόκειται για βίντεο που έχουν κατασκευαστεί και στα οποία το πρόσωπο από το αυθεντικό βίντεο έχει αντικατασταθεί με ένα άλλο πρόσωπο, με τη χρήση τεχνητής νοημοσύνης.<sup>1</sup> Εμφανίστηκαν πρώτη φορά στα τέλη του 2017, όταν ένας ανώνυμος χρήστης της πλατφόρμας Reddit άρχισε να ανεβάζει ψεύτικα πορνογραφικά βίντεο με διάσημους ηθοποιούς, ακολουθώντας τη συγκε-

1. Βλ. άρθρο στην ιστοσελίδα του Vice, «Έκλαιγε και Επέμενε ότι δεν Ήταν στο Βίντεο». Τα Deepfakes Είναι το Επόμενο Υπερόπλο Ψηφιακής Επιρροής;., <https://www.vice.com/el/article/g5gxv3/eklaige-kai-epemene-oti-den-htan-sto-binteo-ta-deepfakes-einai-to-epomeno-yperoplo-pshfiakhs-epirrohs>

κριμένη μεθοδολογία.<sup>2</sup> Ο ανώνυμος αυτός χρήστης έδωσε στα πλαστά βίντεο την ονομασία deepfakes,<sup>3</sup> από τα συνθετικά deep learning, που σημαίνει «βαθιά μάθηση» –δηλαδή τη δυνατότητα ενός μηχανήματος/υπολογιστή να μαθαίνει, βασισμένη στη χρήση τεχνητής νοημοσύνης– και fake, που σημαίνει ψεύτικο, πλαστό.<sup>4</sup>

Με την παραπάνω διαρροή, στην πλατφόρμα του Reddit, του κώδικα που χρησιμοποιείται για τη δημιουργία των deepfakes, εκτινάχθηκε και το ενδιαφέρον των χρηστών του διαδικτύου γι'αυτά, οδηγώντας σε ολοένα αυξανόμενη χρήση και αναπαραγωγή πλαστού περιεχομένου. Αρχικά, στο στόχαστρο των deepfakes βρέθηκαν διάσημα πρόσωπα, όπως ηθοποιοί,<sup>5</sup> τραγουδιστές, αλλά και πολιτικοί.<sup>6</sup> Σήμερα όμως δεν κινδυνεύουν μόνο διάσημα πρόσωπα να αποτελέσουν στόχο των δημιουργών των deepfakes, αλλά και κάθε άλλο πρόσωπο το οποίο χρησιμοποιεί τα μέσα κοινωνικής δικτύωσης και «εκθέτει», υπό μια έννοια, τον εαυτό του, με τη δημοσίευση βίντεο και φωτογραφιών.

Η «συνταγή της επιτυχίας» για την ολοένα αυξανόμενη διάδοση των deepfakes είναι ο συνδυασμός δύο στοιχείων: της αξιοπιστίας και της προσβασιμότητας (believability and accessibility).<sup>7</sup> Όσον αφορά την αξιοπιστία, βασίζεται στο γεγονός ότι οι άνθρωποι έχουμε την τάση να δείχνουμε μεγαλύτερη εμπιστοσύνη σε ό,τι βλέπουμε με τα μάτια μας (ή ακούμε με τα αυτιά μας).<sup>8</sup> Το οπτικό σύστημα του εγκεφάλου μας, όμως, μπορεί εύκολα να ξεγελαστεί, όπως συμβαίνει, για παράδειγμα, με τις οφθαλμαπάτες.<sup>9</sup> Παρ' όλα αυτά, τις περισσότερες φορές, επειδή βλέπουμε κάτι με τα μάτια μας, πιστεύουμε ότι όντως υφίσταται και είναι αληθινό, ακόμα κι αν η λογική μάς λέει ότι κάτι τέτοιο είναι μάλλον απίθανο.

Παράλληλα, το πόσο προσβάσιμα είναι σήμερα πλέον τα εργαλεία για να δημιουργήσει κανείς ένα deepfake βίντεο καθιστά τη διάδοσή τους εξίσου εύκολη και ευρεία. Συγκεκριμένα, τα σύγχρονα τεχνολογικά μέσα και εργαλεία, που συνεχώς αναπτύσσονται, δίνουν τη δυνατότητα στον καθένα να δημιουργήσει πλαστό περιεχόμενο που μοιάζει πολύ με αληθινό, χωρίς να διαθέτει κάποια κατάρτιση ή εξειδικευμένα τεχνικά μέσα.<sup>10</sup>

2. Ο. π.

3. Ορισμένοι ειδικοί τα ονομάζουν αλλιώς «συνθετικά μέσα»

4. Βλ. J. Kietzmann, I.P. McCarthy & L.W. Lee, «Deepfakes: Trick or Treat?», *Business Horizons*, Δεκέμβριος 2019, και S. Ibrahim, “How deepfakes are impacting our vision of reality”, <https://www.swissinfo.ch/eng/how-deepfakes-are-impacting-our-vision-of-reality/46862004>, 2021.

5. No 1 στόχος των δημιουργών deepfakes είναι η Βρετανίδα ηθοποιός Emma Watson, πιθανώς λόγω και της ακτιβιστικής της δράσης.

6. Βλ. J.Kietzmann, I.P. McCarthy & L.W.Lee, ό. π.

7. Ο. π.

8. Βλ. S.Ibrahim, ό. π.

9. Βλ. J. Kietzmann, I.P. McCarthy & L.W.Lee, ό. π.

10. Ο. π.

Συνεπώς, φαίνεται ότι η τεχνολογία για τη δημιουργία πλαστού περιεχομένου αφενός βελτιώνεται συνεχώς και γίνεται πιο αξιόπιστη, αφετέρου καθίσταται προσιτή στα άτομα εξαιρετικά εύκολα, ακόμα και μέσω εφαρμογών που είναι διαθέσιμες στα μέσα κοινωνικής δικτύωσης και επιτρέπουν τη δημιουργία και δημοσίευση τέτοιου είδους περιεχομένου.<sup>11</sup>

Οι επιπτώσεις της δημιουργίας και διάδοσης των deepfakes, καθώς και η επίδρασή τους στην καθημερινή ζωή, είναι ένα ζήτημα που απασχολεί ιδιαίτερος σήμερα τους επιστήμονες, τις κυβερνήσεις, αλλά και την κοινή γνώμη, και που αναμένεται να απασχολήσει ακόμα περισσότερο στο μέλλον. Και αυτό γιατί, όπως όλες σχεδόν οι τεχνολογικές καινοτομίες, έτσι και η τεχνολογία deepfake έχει καλή και κακή πλευρά.

Αυτό που απασχολεί περισσότερο είναι η σκοτεινή πλευρά των deepfakes και οι αρνητικές επιπτώσεις που μπορεί να έχουν στα άτομα και στην κοινωνία. Πρώτα απ' όλα, όσον αφορά τις επιπτώσεις στη δημόσια ζωή, μια τεχνολογία όπως τα deepfakes, που έχουν τη δυνατότητα να κάνουν ένα κατασκευασμένο βίντεο να δείχνει αληθινό, διευκολύνει ιδιαίτερα τη χειραγώγηση της κοινής γνώμης και της ψήφου ανάλογα με συγκεκριμένα συμφέροντα. Για παράδειγμα, βίντεο με ομιλίες ή δηλώσεις πολιτικών προσώπων θα μπορούσαν εύκολα να τροποποιηθούν και να παραποιηθούν, τόσο ως προς το περιεχόμενο των δηλώσεων όσο και ως προς τη συμπεριφορά των εικονιζόμενων προσώπων, με στόχο να πλήξουν την αξιοπιστία τους. Παράλληλα, την εύκολη και γρήγορη διάδοση τέτοιων πλαστών βίντεο θα μπορούσαν να εκμεταλλευτούν οι πολιτικοί τους αντίπαλοι, προς ίδιον όφελος και για δική τους ανάδειξη. Χαρακτηριστικό είναι το περιστατικό με τον Πρόεδρο της Γκαμπόν Ali Bongo, ο οποίος, μετά από αρκετούς μήνες απουσίας από τη δημόσια ζωή, εμφανίστηκε σε ένα βίντεο για να απευθύνει ευχές για το νέο έτος την Πρωτοχρονιά. Αρκετοί που παρακολούθησαν το βίντεο στα μέσα κοινωνικής δικτύωσης παρατήρησαν πως ήταν ασυνήθιστη η συμπεριφορά του Προέδρου Bongo, με αποτέλεσμα να ισχυριστούν πως πρόκειται για deepfake και να τροφοδοτήσουν φήμες περί σοβαρής ασθένειας ή θανάτου του. Οι φήμες αυτές οδήγησαν σε απόπειρα πραξικοπήματος από το Στρατό, η οποία όμως απέτυχε, ενώ ο Πρόεδρος Bongo μετά το περιστατικό επανήλθε κανονικά στις δημόσιες εμφανίσεις του.<sup>12</sup>

Σημαντικός είναι ο ρόλος των deepfakes και σε οικονομικές απάτες. Χαρακτηριστική είναι η περίπτωση ενός Βρετανού διευθυντή επιχείρησης (CEO) ο οποίος εξαπατήθηκε μέσω audio deepfake, και θεωρώντας ότι έχει μιλήσει με τον διευθυντή συγγερικής εταιρείας, ενέκρινε την αποστολή 250.000 λιρών, δηλαδή σχεδόν 287.000 ευρώ.<sup>13</sup> Ταυτόχρονα, η τεχνολογία αυτή

11. Ο. π.

12. Βλ. Henry Ajder, Giorgio Patrini, Francesco Cavalli & Laurence Cullen, *The State of Deepfakes: Landscape, Threats, and Impact*, [https://regmedia.co.uk/2019/10/08/deepfake\\_report.pdf](https://regmedia.co.uk/2019/10/08/deepfake_report.pdf), καθώς και το προαναφερθέν άρθρο στην ιστοσελίδα του Vice.

13. Βλ. το προαναφερθέν άρθρο στην ιστοσελίδα του Vice.

επιτρέπει την ανάπτυξη ενός νέου είδους εκβιασμών, των τεχνολογικών εκβιασμών, όπου στόχος των εκβιαστών μπορεί να είναι είτε η αποκόμιση οικονομικού οφέλους είτε η αποτροπή οποιασδήποτε άλλης ενέργειας από το θύμα του εκβιασμού.<sup>14</sup>

Ο πιο σπουδαίος ρόλος, όμως, της τεχνολογίας των deepfakes είναι αυτός που διαδραματίζουν στην άνοδο της εκδικητικής πορνογραφίας. Ο όρος «εκδικητική πορνογραφία» χρησιμοποιείται για να περιγράψει το φαινόμενο της δημοσιοποίησης φωτογραφιών ή βίντεο σεξουαλικού περιεχομένου χωρίς τη συναίνεση του εικονιζόμενου προσώπου.<sup>15</sup> Τα δεδομένα που υπάρχουν σχετικά με τη χρήση των deepfakes στην εκδικητική πορνογραφία είναι τουλάχιστον ανησυχητικά. Η εύκολη πρόσβαση σε εργαλεία για την κατασκευή ενός πλαστού βίντεο σεξουαλικού περιεχομένου, όπου το εικονιζόμενο πρόσωπο δεν έχει καν συμμετάσχει, ούτε και συναινέσει στη δημιουργία του, καθιστά ακόμα μεγαλύτερο τον κίνδυνο θυματοποίησης οποιουδήποτε προσώπου. Σήμερα, σύμφωνα με πρόσφατα δεδομένα, το 96% των deepfake βίντεο που υπάρχουν στο διαδίκτυο είναι σεξουαλικού περιεχομένου.<sup>16</sup> Σε αυτά, κατ' αποκλειστικότητα, σε ποσοστό 100%, τα θύματα είναι γυναίκες.<sup>17</sup> Σε ορισμένες από αυτές τις περιπτώσεις, βέβαια, τα κίνητρα μπορεί να μην είναι εκδικητικά αλλά οικονομικά: ο δράστης κατασκευάζει ψεύτικο πορνογραφικό υλικό και στη συνέχεια εκβιάζει το θύμα, ζητώντας χρήματα, προκειμένου να μην το δημοσιεύσει.<sup>18</sup> Όποια κι αν είναι όμως τα κίνητρα, φαίνεται ότι βρισκόμαστε μπροστά σε μια νέα απειλή, όπου ο καθένας μπορεί να πέσει θύμα έκθεσης μέσω δημοσίευσης πλαστών βίντεο σεξουαλικού περιεχομένου, στη δημιουργία των οποίων δεν έχει καν συναινέσει ή συμμετάσχει. Οι αρνητικές επιπτώσεις αυτής της έκθεσης για το θύμα είναι εξαιρετικά σοβαρές.<sup>19</sup> Το άγχος, ο φόβος και η ντροπή που βιώνει αποτελούν ένα είδος ψυχολογικής βίας, ενώ η καθημερινότητα και οι διαπροσωπικές σχέσεις του θύματος αρκετές φορές διαταράσσονται ανεπανόρθωτα.

Από όσα αναπτύχθηκαν παραπάνω, φαίνεται ότι η τεχνολογία των deepfakes δημιουργεί νέες «προοπτικές» για το έγκλημα, καθώς εισάγει νέες μεθόδους και τεχνικές εγκληματικής δράσης.

14. Βλ. για παράδειγμα την περίπτωση της Ινδής δημοσιογράφου Rana Ayyub, η οποία βρέθηκε να πρωταγωνιστεί σε βίντεο σεξουαλικού περιεχομένου, η δημοσίευση των οποίων είχε ως στόχο τον εκβιασμό της προκειμένου να σταματήσει την έρευνα για τον βιασμό και τη δολοφονία ενός οκτάχρονου κοριτσιού από το Κασμί (από το προαναφερθέν άρθρο στην ιστοσελίδα του Vice).

15. Αλλιώς «μη συναινετική πορνογραφία» και «σεξουαλική κακοποίηση μέσω εικόνας». Βλ. Αγγελική Γιαννάκη, «Η Μη Συναινετική Πορνογραφία στην Ελληνική Έννομη Τάξη», *The Art of Crime*, Μάιος 2021.

16. Βλ. Ajder, Patrini, Cavalli & Cullen, ό. π.

17. Ο. π..

18. Πρόκειται για το λεγόμενο sextortion. Βλ. το προαναφερθέν άρθρο στην ιστοσελίδα του Vice.

19. Βλ. Chesney, Robert & Citron, Danielle Keats, «Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security» (14 Ιουλίου 2018), 107 *California Law Review* 1753 (2019), *U of Texas Law, Public Law Research Paper* No. 692, *U of Maryland Legal Studies Research Paper* No. 2018-21, <https://ssrn.com/abstract=3213954> or <http://dx.doi.org/10.2139/ssrn.3213954>.

Η εγκληματική δράση με τη χρήση τεχνολογίας deepfake εμπίπτει στην κατηγορία των εγκλημάτων *με* χρήση τεχνητής νοημοσύνης (crimes *with* Artificial Intelligence), όπου η τεχνητή νοημοσύνη αποτελεί το εργαλείο για τη διάπραξη του εγκλήματος.<sup>20</sup>

Δεν είναι βέβαιο αν τελικά τα deepfakes οδηγούν σε αύξηση της εγκληματικότητας. Είναι, όμως, πολύ πιθανό να συμβάλλουν στην αύξηση του φόβου του εγκλήματος. Ο φόβος του εγκλήματος δεν συνδέεται μόνο με την πιθανότητα θυματοποίησης, αλλά και με την πρόσληψη της κοινωνικής πραγματικότητας και τις υφιστάμενες δυνατότητες αντίδρασης και προστασίας.<sup>21</sup> Είναι σαφές ότι η τεχνολογία των deepfakes έχει εξελιχθεί τόσο ώστε μπορεί να ξεγελάσει όχι μόνο την ανθρώπινη όραση, αλλά και τους αλγόριθμους που έχουν δημιουργηθεί για να ανιχνεύουν πλαστό περιεχόμενο. Για τους επιστήμονες, αποτελεί πραγματικό αγώνα δρόμου ενάντια στο χρόνο και την εξέλιξη της τεχνολογίας η προσπάθεια να δημιουργήσουν αξιόπιστα εργαλεία τα οποία θα ανιχνεύουν επιτυχώς τα deepfakes και θα παρέχουν ασφάλεια σε όλους.<sup>22</sup> Αυτή η αβεβαιότητα, όμως, σχετικά με τις δυνατότητες αντίδρασης και προστασίας που υφίστανται σήμερα απέναντι στην τεχνολογία των deepfakes μπορεί να οδηγήσει σε αύξηση του φόβου του εγκλήματος, καθώς ο καθένας θα θεωρεί τον εαυτό του εκτεθειμένο στον κίνδυνο της θυματοποίησης, ενώ παράλληλα στην καθημερινότητά του δύσκολα θα μπορεί να εμπιστευτεί αυτά που βλέπει ή ακούει.

Έτσι, παραμένει το ερώτημα σχετικά με τις αρνητικές επιπτώσεις της χρήσης και διάδοσης των deepfakes, και αν αυτές συνδέονται με πραγματική αύξηση των δεικτών της εγκληματικότητας ή με την ανασφάλεια και την αβεβαιότητα που προκαλούν στον μέσο άνθρωπο.

Καταλήγοντας, όπως η τεχνητή νοημοσύνη γενικά, έτσι και η τεχνολογία των deepfakes δεν μπορεί να είναι από μόνη της βλαβερή. Αντιθέτως, μπορεί να είναι χρήσιμη σε αρκετές περιπτώσεις. Σήμερα, είναι ήδη γνωστή η χρήση των deepfakes στην ψυχοθεραπεία, στο πλαίσιο της προσπάθειας να μειωθεί ο πόνος που προκαλεί η απώλεια ενός αγαπημένου προσώπου. Ένα τέτοιο εργαλείο είναι το DeepNostalgia, που χρησιμοποιείται από την πλατφόρμα MyHeritage και μέσω του οποίου ο χρήστης μπορεί να δημιουργήσει κινούμενες εικόνες με αγαπημένα πρόσωπα που έχει χάσει.<sup>23</sup>

20. Άλλες κατηγορίες εγκλημάτων που αναφέρονται είναι τα «crimes on Artificial Intelligence» και «crimes by Artificial Intelligence». Βλ. K.J. Hayward & M.M. Maas, «Artificial Intelligence and Crime: A primer for Criminologists», <https://journals.sagepub.com/doi/10.1177/1741659020917434>, 2020.

21. Βλ. Χ.Ζαραφονίτου, «Φόβος του Εγκλήματος», στο Κ. Δ. Σπινέλλη, Ν. Ε. Κουράκης και Μ. Π. Κρανιδιώτη (επιμ.), *Λεξικό Εγκληματολογίας*, Τόπος 2018.

22. Βλ. S.Ibrahim, ό. π.

23. Ό. π.