„Update Verfügbar – ein Podcast des BSI“

Transkription für Folge 22, 29.07.2022:
Deepfakes – die perfekte Täuschung?

Moderation: Ute Lange, Michael Münz
Gast: Markus Ullmann, BSI
Herausgeber: Bundesamt für Sicherheit in der Informationstechnik (BSI)

Lange: Hallo und herzlich Willkommen zu einer neuen Folge von „Update Verfügbar“, dem Podcast für Sicherheit im digitalen Alltag. Mein Name ist Ute Lange.

***Münz**:* Ich bin Michael Münz. In dieser Folge geht es um ein Thema, das gerade intensiv diskutiert wird, und zwar Deepfakes. Was hat es mit diesen Fälschungen auf sich? Wann sind sie deep und wann eher shallow? Warum ist es wichtig, diese im digitalen Alltag zu erkennen? Darüber sprechen wir mit Markus Ullmann vom Bundesamt für Sicherheit in der Informationstechnik. Ich bin bereits sehr neugierig. Es gibt vieles zu besprechen und ich bin sicher, dass es ein weiteres, erkenntnisreiches Gespräch wird. Letzte Folge hatten wir auch eins. Ute, du klangst am Ende der Folge ein bisschen nach Information-Overload. Konntest du die Folge inzwischen verarbeiten?

Lange: Ja. Letztes Mal hatten wir Immanuel und Tim zu Gast, zwei Hacker, die auf der guten Seite der Macht stehen. Sie nutzen ihr Wissen über die Hintergründe von vielen der Angriffe, die wir besprochen haben, für das Allgemeinwohl. Das war einer der Aspekte, die für mich hängengeblieben sind, und zwar dass sie die Motivation haben, Fähigkeiten, mit denen Hackerinnen und Hacker destruktiv arbeiten, für das andere zu nutzen, nämlich um Dinge zu erkennen, nachzuvollziehen, was bei Angriffen oder Hacks passiert, und aufzuklären. Nicht nur bei uns im Podcast, sondern in ihrem Umfeld, bei ihren Kunden sowie im Bekanntenkreis, sodass wir alle besser geschützt sind. Das ist bei mir hängengeblieben. Ich konnte mir auch gut merken, dass Immanuel ziemlich am Ende der Folge „Kopf und Kontext“ gesagt hatte. Beispielsweise, passt die E-Mail oder die SMS, die ich jetzt lese, in meinen Kontext? Ist das eine Firma, bei der ich bestelle? Ist das eine Bank, bei der ich ein Konto habe? Ist vielleicht die Person mir bekannt? Der Kopf bedeutet, ob mein Kopf gerade klar ist und ob ich aufmerksam für das bin, was ich lese. Oder bin ich wieder mit 35 Sachen gleichzeitig beschäftigt und sollte lieber ein Stück zurücktreten und mir das in Ruhe noch mal angucken, bevor ich irgendwas mache, was nachher zu einem Problem führt? Das sind meine zwei Takeaways, die ich gut in Erinnerung behalten habe. Und bei dir? Ist dir nach dem Gespräch vielleicht ein einsamer USB-Stick begegnet? Wenn ja, hast du ihn mitgenommen und eventuell sogar irgendwo eingesetzt?

***Münz**:* Selbst, wenn ich einen gesehen hätte, hätte ich ihn natürlich nicht mitgenommen. Das Thema „Faktor Mensch“ hat mich sehr bewegt, und zwar wie man Menschen dazu bewegen kann, unvorsichtig zu sein. Es reicht am Ende, wenn eine Mitarbeiterin oder ein Mitarbeiter unvorsichtig ist und auf einen Link klickt, einen unbekannten USB-Stick einsteckt oder ein Passwort weitergibt, damit der Zugang zum Netzwerk da ist. Ich habe dieses Menschen manipulieren, damit die etwas machen, was mir in meinen Plan passt, wirklich als wichtig abgespeichert.

Lange: Man nennt das Social Engineering, also das Nutzen unserer Gefühle und Bedürfnisse, wie beispielsweise Angst oder Neugierde, auf der Suche nach einem Schnäppchen. Das BSI hat kürzlich vor einem Fall im Bundestag gewarnt, bei dem Kriminelle die Zielpersonen, in dem Fall Abgeordnete oder deren Mitarbeiterinnen und Mitarbeiter, kontaktierten. Sie haben eine SMS von Kriminellen bekommen, die sich als Politikerinnen und Politiker aus einem ähnlichen Umfeld ausgaben. Man wolle gerne Kontakt aufnehmen und ein vertrauliches Gespräch führen. Dafür solle die angesprochene Person auf einen sicheren Messenger Dienst wie Signal, WhatsApp oder Telegram wechseln, damit das erfolgen kann.

***Münz**:* Komisch ist der Wechsel auf ein anderes System, wenn der Kontakt sowieso schon da ist.

Lange: Es soll ja vertraulich sein. Man vermittelt den Eindruck, dass keiner mithören kann, wenn man auf ein anderes Medium geht. Auf diesem Weg sollen Betroffene dazu gebracht werden, sich bei einem der Dienste ein neues Konto anzulegen. Dann versuchen die Täterinnen und Täter unter einem Vorwand den Authentifizierungscode zu erhalten, der genutzt werden kann, um im Namen der Person, die man angesprochen hat, ein weiteres Konto einzurichten. Die Motive sind noch nicht klar, aber sobald jemand deinen Authentifizierungscode hat, könnte er in deinem Namen ein weiteres Konto einrichten und vielleicht über den Messenger Dienst als dich ausgebend mit Personen in Kontakt treten. Das BSI sagt, dass das nichts Neues ist, unter anderem weil man dafür nicht unbedingt viel technisches Know-how braucht, aber sie wollen jetzt noch mal sensibilisieren. Was das Interessante oder Mysteriöse war, in derselben Woche, als ich das gelesen habe, ist bei mir in einem dieser Dienste aufgepoppt, dass eine gute Freundin von mir jetzt diesen Messenger Dienst nutzt. Da war aber ein seltsames Foto, und zwar nicht ein Foto von ihr, sondern von einem Mann. Und es war auch nicht ihre Telefonnummer. Ich habe sie auf einem anderen Kanal angerufen und gefragt, ob sie diesen Messenger hat. Sie sagte, dass sie die App schon vor ein paar Wochen gelöscht hat und da keinen Account mehr hat. Ich habe gedacht, noch mal den Tipp an alle herausgeben: Wenn Personen auf einem Messenger Dienst ein neues Konto haben, das einem allerdings fishy vorkommt, frag sie lieber auf einem anderen Weg, ob sie das tatsächlich sind. In diesem Fall beispielsweise war es niemand, den ich kannte. Das gilt vielleicht auch für die Politikerinnen und Politiker, die von dieser Masche offensichtlich betroffen sind.

***Münz**:* Offensichtlich ist eine gewisse Skepsis heutzutage nicht nur bei E-Mails oder Messenger Diensten, sondern auch bei Telefonanrufen echt angebracht. Die

Bundesnetzagentur verzeichnet nämlich einen erheblichen Anstieg von Beschwerden über die Fake-Anrufe vermeintlicher Polizeibehörden. Die Bundesnetzagentur erhält Hinweise darauf, dass Verbraucherinnen und Verbraucher Anrufe im Namen von Europol oder anderen vermeintlichen internationalen Polizeibehörden wie Interpol oder FBI erhalten, und hat im Juni mehr als 7600 Beschwerden verzeichnet. Im Vergleich dazu ging die Zahl zu Anfang des Jahres gegen null. Da ist gerade richtig viel los.

Lange: Welche Masche verbirgt sich dahinter? Wer ist die Zielgruppe und was soll damit erreicht werden?

***Münz**:* Ziel der Anrufer sind offensichtlich ältere Menschen. Das Telefon klingelt und bei Annahme des Anrufers ist eine Bandansage zu hören. Man wird aufgefordert, auf dem Tastenfeld die eins zu drücken, um weitergeleitet zu werden. Dieses Vorgehen kennt man aus Warteschlangen. Nach dem Drücken der Taste erfolgt die Weiterleitung zu einer Person, die teils in englischer Sprache zur Herausgabe persönlicher Daten oder zur Überweisung von Geld auffordert. Fies ist die Tatsache, dass die Absenderkennung der Anrufe deutsche Rufnummern sind, weil da etwas manipuliert worden ist. Eigentlich haben die Anrufe ihren Ursprung im Ausland. Die Bundesnetzagentur warnt davor, den Anrufenden persönliche Daten mitzuteilen oder Geldzahlung zu leisten. Stattdessen sollte man den Anruf beenden, Taste „ignorieren“ drücken und, damit die Tatsache, dass so etwas passiert, dokumentiert wird und solche Anrufe erfasst werden können, Anzeige bei der örtlichen Polizeidienststelle stellen.

Lange: Hier sprechen wir wieder das Thema Faktor Mensch an und jetzt offensichtlich mit Angst. Viele reagieren vielleicht darauf, dass eine Polizeibehörde offiziell anruft und etwas von einem will. Das heißt, Skepsis soll nicht nur da sein, wenn die Messenger Dienste neue Personen angeben, sondern auch am Telefon. Skepsis ist ein gutes Stichwort für unser Thema heute. Wir möchten über Deepfakes sprechen und haben Markus Ullmann vom BSI eingeladen. Markus, wir freuen uns sehr, dass du da bist! Stell dich bitte kurz vor. Was ist dein Hintergrund? Wie bist du ins BSI gekommen und was machst du heute dort?

Ullmann: Hallo Ute und Michael. Danke, dass ich heute bei euch sein darf. Gerne sage ich was dazu. Ich bin ursprünglich Elektroingenieur. Anfang 1991, als das BSI gegründet wurde, wurden Kolleginnen und Kollegen gesucht. IT-Sicherheit war für mich vollkommen neu und spannend und ich bin immer auf der Suche nach spannenden Sachen. So habe ich mich beim BSI beworben und man hat mich genommen. Es war nicht geplant und ich hatte mir gar nicht vorgestellt, dass ich so lange beim BSI bleibe, aber ich fühle mich dort sehr wohl und habe seit mittlerweile über 30 Jahren mit sehr spannenden Themen zu tun. Mittlerweile darf ich auch eine Gruppe beziehungsweise ein sogenanntes Referat leiten, wo wir Themen von Blockchain, digitales Zentralbankgeld, autonomes Fahren, Angriffe auf autonome Fahrfunktion bis hin zu biometrischen Systemen und Deepfakes haben. Wir gestalten durch Angriffe und Kaputtmachen, das ist immer unser Einstieg. Es ist kein Selbstzweck, sondern wir wollen die Systeme und ihre Schwachstellen verstehen, Anforderungen und Prüfkriterien definieren, damit wir am Ende des Tages zu besseren und sicheren Systemen kommen. Das

ist immer unser Regelkreis: erst kaputt machen und dann schauen, wie man es besser machen kann.

Münz: Bevor wir ins Thema einsteigen und uns intensiver über Deepfakes und vielleicht biometrische Systeme austauschen, haben wir für dich die obligatorische Entweder-oder-Frage. Du hast gerade das Stichwort 30 Jahre genannt. Wir wollen von dir gerne wissen, wenn du die Möglichkeit hättest, würdest du die 30 Jahre zurückreisen und da etwas anderes machen oder lieber 30 Jahre nach vorne in die Zukunft reisen?

Ullmann*:* Ich denke, ich würde eher 30 Jahre in die Zukunft reisen, weil mich immer spannende neue Themen interessiert haben. Ich durfte sicherlich viele interessante Sachen in der Vergangenheit machen, aber fände die Zukunft spannender. Man kann auch alten Kram wieder aufwärmen, aber das wäre eher nicht so mein Ding.

Lange: Danke dir. Wir bleiben jetzt in der Gegenwart, weil das Thema Deepfakes einige Aufmerksamkeit in den Medien erregt hat und vielleicht bei dem einen oder anderen Zuhörerin oder Zuhörer Interesse geweckt hat. Es kam kürzlich raus, dass mehrere Bürgermeisterinnen und Bürgermeister in europäischen Hauptstädten mit einem vermeintlichen Vitali Klitschko, ihrem Amtskollegen aus Kiew, zu einem Videocall verabredet waren. Es stellte sich nach und nach heraus, dass Klitschko in all diesen Gesprächen an dem anderen Ende der Leitung eine Fälschung war. Es hieß, es sei ein Deepfake. Kannst du mit deinen Kenntnissen ein bisschen beleuchten, was da genau passiert ist?

Ullmann: Ich versuche das. Zu Beginn möchte ich aber sagen, dass wir selbst keine Quellinformationen haben, die wir uns anschauen konnten, sondern ich kann nur das berichten, was wir über die Medien mitbekommen haben, insbesondere vom Investigativjournalisten Daniel Laufer, der den Fall schlüssig analysiert hat. Wir haben verstanden, dass die Staatskanzlei für ein Interview angefragt worden ist, wobei das Thema sicherer Kanal eine große Rolle spielt. Das heißt, wer hat eigentlich angefragt? War die Annahme, dass es aus dem Klitschko-Umfeld kommt, richtig? Das war sie am Ende nicht. Wir wissen, dass Herr Klitschko mit dem ukrainischen Journalisten Dmitry Gordon im April einen Videocall durchgeführt hat. Dieser Call ist öffentlich gestellt worden und ist verfügbar. ARD-Journalist hat von der Staatskanzlei fünf Fotos von dem Gespräch bekommen. Das Gespräch wurde nicht komplett aufgezeichnet, sondern es wurden angeblich nur fünf Fotos gemacht, quasi Momentaufnahmen aus diesem Videocall. Der Journalist konnte nachweisen, dass alle fünf Fotos inklusive Kontext in dem Video aus dem April drinstecken. Es sind keine Veränderungen gemacht worden, sodass die große Annahme jetzt ist, dass man sich das alte Videocall vom April hergenommen hat und Frames rausgeschnitten hat, die offline zusammengefügt und nachher abgespielt wurden. Wenn man vielleicht noch irgendwelche Bildstörungen mit einbaut, die einen schlechten Kanal simulieren sollen, ist es bereits die erste Technik. Ich kann leider nicht viel zur Audiospur und dem, was eigentlich passiert ist, sagen. Ich habe gehört, dass ein Übersetzer mit eingebunden war, obwohl ich das nicht genau weiß. Das macht aber deutlich, wieso die Synchronisierung zwischen Mundbewegungen und der Sprachausgabe nicht als ein möglicher Fake aufgefallen ist. In

einem normalen Video muss die Mundbewegung offensichtlich zur Sprache passen. Allerdings wenn ein Zwischenhändler eingeschaltet wird, ist es von vornherein klar, dass es nicht synchron sein muss oder kann, weil ein anderer spricht. Das war von da gesehen kein Deepfake, sondern eine andere Manipulation. Ich glaube aber, dass das Entscheidende nicht die Frage ist, ob es ein Deepfake, oder ein anderer Fake war, sondern die Tatsache, dass es einfach ein Fake war und die Bürgermeisterin, die Staatskanzlei in Berlin oder die Senatskanzlei darauf reingefallen sind.

Münz: Es ist natürlich interessant zu hören, wie das technisch gemacht worden ist. Es klingt aber eigentlich so, als wäre das eigentlich ein Beispiel für Social Engineering, also wie Menschen über bestimmte Methoden und Kommunikationswege dazu bewegt worden sind, den Termin anzunehmen, und wie wir mittlerweile wissen, nicht nur in Berlin, sondern auch in anderen Städten.

Ullmann*:* Genau. Diese Komiker, Vovan und Lexus, die für eine Gazprom-Tochter arbeiten und dafür gut bezahlt worden sind, scheinen sich zum Ziel zu machen, westliche Politikerinnen und Politiker auf den Arm zu nehmen. Das war jetzt ein Beispiel.

Lange: Was sind für dich mit deinem Hintergrundwissen Lehren daraus? Worauf sollte eine Senatskanzlei oder ein Bürgermeisterbüro achten, wenn solche Anfragen kommen? Vielleicht war es in diesem Fall der Promi-Faktor? Wer möchte in der heutigen Zeit nicht mit Klitschko sprechen? Was könnten Dinge sein, die man jetzt daraus lernen kann? Nicht nur technisch, sondern für das eigene Schutzverhalten.

Ullmann*:* Ihr habt das ganz zu Beginn angesprochen, und zwar überhaupt zu prüfen, ob das eine valide Anfrage ist und ob die Anfrage in diesem Fall wirklich aus dem Umfeld von Vitali Klitschko kommt. Vielleicht könnte man auf einem zweiten Kanal verifizieren und rückfragen, ob sie das wirklich sind. Das ist meiner Meinung nach der springende Punkt, und zwar nach Möglichkeit im Vorfeld zu klären, ob man mit einer vertrauenswürdigen Quelle spricht oder ob man jemandem aufsetzt. Man geht immer vom guten Fall aus, aber wenn sowas jetzt vor allem im politischen Raum gezielt gemacht wird, sollte man noch mal die Verifikationsschleife ziehen und das einfach im Vorfeld checken. Der Schaden kann beträchtlich sein, wenn man auf den Leim geführt wird.

Münz: Über die Schäden, die durch solche Geschichten auftreten können, würden wir gerne gleich mit dir sprechen. Zuerst würde ich gerne von dir wissen, was du gedacht hast, als du mit deinem Hintergrund gehört hast, dass es ein Deepfake hätte sein können. Hast du gedacht, dass sowas gut sein kann oder könntest du dir nicht vorstellen, dass das jemand so auf die Kette kriegt?

Ullmann: Ich muss ehrlich sagen, dass ich mir das nicht genau angeschaut habe. Es ist schon in der Vergangenheit gelaufen und ich gucke lieber mehr in die Zukunft. Erst als die Anfrage von euch kam, habe ich mit Kolleginnen und Kollegen ein bisschen näher hingeschaut, was passiert ist. Ich würde auf keinen Fall ausschließen, dass so etwas grundsätzlich nicht als Deepfake passieren könnte. Wir sind im Umfeld von Deepfakes praktisch unterwegs, sowohl

was die Videospur angeht als auch die Möglichkeit, synthetisch Audio für eine Zielperson zu adaptieren. Das heißt, ich spreche so wie ich jetzt spreche, aber klinge wie du, Ute, oder du klingst wie Michael. Unsere Aufgabe ist es, den Angriff oder sein Bedrohungspotenzial zu erfassen. Was geht mit Tools, die wir nutzen? Sollen wir nur öffentlich verfügbare Tools nutzen und im nächsten Schritt klären oder analysieren, was man heute tun kann, um so etwas automatisiert zu erkennen? Da muss es hingehen. Genau deswegen könnte ich mir vorstellen, dass das ein Deepfake gewesen sein könnte, was sich im Nachhinein als nicht so herausgestellt hat, aber es wäre grundsätzlich möglich.

***Lange**:* In dem Kontext sind eine ganze Menge Begriffe formuliert worden, unter anderem Deepfake, Shallowfake, Cheapfake etc. Kannst du uns von der technologischen Seite erklären, wo die Unterschiede liegen?

Ullmann: Ich fange damit an, über was wir gerade gesprochen haben, und zwar Vitali Klitschko. Man hat sehr wahrscheinlich ein altes, öffentliches Video genommen, aus diesem Video Sequenzen rausgenommen und die geschickt zusammengebaut. Es ist wahrscheinlich keine Künstliche Intelligenz eingesetzt worden. Dann wäre es ein einfacher Fake. Deepfake ist immer irgendetwas, wobei man tiefe und neuronale Netze einsetzt, „Deep“ steht für tief. Man trainiert das Funktionsverhalten dieser tiefen neuronalen Netze für eine Aufgabe. Das Stichwort ist maschinelles Lernen. Deepfake bedeutet, dass man tiefe neuronale Netze verwendet hat, um den Angriff durchzuführen. Wenn wir jetzt beispielsweise für dich, Ute, einen Audio-Deepfake machen wollten, bräuchten wir eine möglichst lange Sprachausgabe von dir. Dann würden wir ein vortrainiertes Audiosystem mit deinen Sprachsamples auf deine Person trainieren. Es gibt zwei Klassen von Systemen. Die einen sind Voice-Conversion-Systeme, die es dir, Michael, ermöglichen würden, zu sprechen und so wie Ute klingen, wenn das Zielsystem auf sie trainiert wäre. Die anderen sind Text-to-Voice-Systeme. Das heißt, dass ich dem System einen Text gebe, und wenn das System auf dich trainiert ist, würde es aus dem Text eine synthetische Sprachausgabe für deine Person machen. Wir beobachten das seit etwa zwei Jahren, im Wesentlichen frei verfügbare Tools, die man frei nutzen kann. Ich muss sagen, es was vor gut einem Jahr noch sehr schwierig, Sprachausgabe so zu gestalten, dass Emotionen enthalten seien. Sie klangen sehr monoton. Das ist über das letzte Jahr deutlich besser geworden, weil die technische Entwicklung vorangeschritten ist. Wenn man aber in die Richtung Gefahrenpotenzial schaut, muss man sagen, dass es zur realen Gefahr werden kann, weil es immer authentischer klingt. Die Gefahr für jemanden, der nicht unterscheiden kann, ob das eine wahre oder eine synthetisch erzeugte Stimme ist.

Münz: Bevor wir ins Thema Sprache einsteigen, weil das in fast allen alltäglichen Anwendungen, wie beispielsweise Sprachassistenten, zum Tragen kommt, würde ich gerne noch mal versuchen, das Thema Deepfakes beziehungsweise Manipulation von Bild und Ton zu kategorisieren. In dem Klitschko-Fall sehen wir, dass man vorhandenes Material in der neuen Kombination zusammengeschnitten hat und eine Audiospur draufgelegt hat, um die Spuren des Bestehenden zu verwischen. Wir kennen außerdem die Bahnansagen, die so

klingen, als hätte man vorgefertigte Elemente wie „der Zug fährt vom Gleis sieben um 17 Uhr und hat Verspätung“, die neu zusammengefügt werden und eher emotionslos sind. Die Königsklasse ist ein System, das im technisch herausragendsten Fall spontan auf Fragen und Input reagieren könnte, sowie auf Gesprächsinhalte, die in einer Situation auftreten.

Ullmann: Ich glaube, das wäre noch einen Schritt weiter. Ich habe versucht zu erläutern, dass es heute frei verfügbare Systeme, die sogenannten Sprachausgabesysteme, gibt, die man auf eine Zielperson trainieren kann. Das heißt, ich spreche in das Mikrofon rein, danach wird eine Voice-Conversion gemacht. Diese Sprache wird in das tiefe neuronale Netz gefüttert und man bekommt eine Tonausgabe mit Sprache von Ute.

Lange: Heißt das, dass du mir Dinge in den Mund legen könntest, die ich vielleicht niemals sagen würde oder die für mich nicht unbedingt mit einer positiven Resonanz belegt wären?

Ullmann*:* Wenn das mein Ziel wäre, könnte ich das tun.

Lange: Du hast gesagt, es ist im letzten Jahr bei dieser Sprachsystem-Technologie ein Sprung passiert. Wer verfügt zur heutigen Zeit über die Fähigkeiten oder die Tools, die nicht frei zugänglich oder kostengünstig sind, um so etwas herstellen zu können? Gibt es schon etwas oder ist es Zukunftsmusik und wir sind alle nur bei dem Begriff immer gleich alarmiert?

Ullmann: Nein, es gibt bereits eine Reihe von Open Source Tools, die man verwenden kann. Wir arbeiten auch mit solchen Open Source Tools und verbessern die mit unseren Möglichkeiten. Wir haben vor anderthalb Jahren angefangen, in diese Technologie einzusteigen, die mit sowohl Video- als auch Audiomanipulation zu tun hat. Obwohl wir IT-Sicherheitsexperten für alle möglichen Bereiche sind, hatten wir bisher keine Berührung mit dieser Technologie und sind quasi erst eingestiegen. Heute sind wir schon gut unterwegs. Das heißt, jemand, der die Zeit aufbringt und die Ressourcen dafür hat, könnte das vergleichbar wie wir tun. Die Tools, die man grundlegend dafür braucht, sind öffentlich verfügbar, es gibt außerdem Forschungscommunities, die diese Thematik weiter vorantreiben. Es ist grundsätzlich nicht schlecht. Filmindustrie wäre ein Beispiel dafür. Der Schauspieler oder die Schauspielerin, der oder die gerade beim Filmdreh eingesetzt werden soll, hat beispielsweise eine Covid-Erkrankung und kann nicht sprechen.Die Filmcrew möchte den Film aber trotzdem weiterbearbeiten. Das wäre ein Beispiel für legale Einsatzmöglichkeit der Technologie. Interessant wird es immer dann, wenn sie missbraucht wird. Das kann man mit allem machen, selbst mit einem Brotmesser zu Hause, mit dem man normalerweise Brot schneidet – auch das könnte man anders Verwenden als ursprünglich gedacht, das ist hier vergleichbar.

Münz: Hast du vielleicht ein konkretes Beispiel dafür, wie ein missbräuchlicher Einsatz zu Schäden geführt hat, beziehungsweise wie es wirklich zu kriminellen Aktivitäten gekommen ist? Beim Thema Sprache fällt mir spontan der CEO Fraud ein. Der Chef oder die Chefin einer Firma möchte von einer Mitarbeiterin oder einem Mitarbeiter gerne dies und jenes gemacht bekommen. Früher gab es E-Mails mit der Bitte an Kolleginnen und Kollegen, beispielweise 27.000 € auf ein bestimmtes Konto zu überweisen. Heute könnte bereits die Sprache beim

Telefonanruf so gut gefälscht werden, dass man nicht merkt, dass man nicht mit seinem Chef gesprochen hat. Letztendlich stellt sich heraus, dass es nicht der Fall ist. Wir kennen nur zwei Fälle, von denen wir aus der Presse erfahren haben. Einmal war es ein CEO Fraud, der in einem englischen Unternehmen passiert ist. Der zweite Fall, der bekannt geworden ist, war der Fall eines Hongkonger Bankdirektors. Wie gesagt, wie haben das aus der Presse erfahren. Ob das wirklich stimmt, muss ich offenlassen.

Lange: Ich würde kurz auf eine Schlagzeile, die ich im Nachgang zum Klitschko-Fall gelesen habe, zurückkommen. Es stand, dass die Angst vor Deepfakes gefährlicher als Deepfakes selbst ist oder sein könnte. Mir ist der Fall in Gabun eingefallen, über den wir in der ersten Folge gesprochen haben und der vor ein paar Jahren stattgefunden hat. Der Staatspräsident war länger krank und nicht in der Öffentlichkeit, während es wegen politischer Opposition Unruhe im Land gab. Der Präsident hat eine Ansprache gehalten und wirkte dadurch, dass er einen Schlaganfall hatte, sehr mechanisch. Manche Akteurinnen und Akteure, unter anderem das Militär, haben die Situation für die Behauptung ausgenutzt, dass hier etwas kriminelles vorgeht und der Präsident schon lange tot sei. Es hat zu Unruhen geführt, die aber nach einigen Tagen beruhigt werden konnten. In diesem Fall hat die Angst vor Deepfakes oder die Behauptung ohne Belege, dass es ein Fake sei, zu politischen Schäden geführt. Wie siehst du diese Überschrift, und zwar dass unsere Angst vor Deepfakes viel gefährlicher als Deepfakes selbst ist?

Ullmann*:* Ich denke, dass Angst generell ein schlechter Begleiter ist. Wie ihr zu Beginn des Gesprächs erwähnt habt, muss man einfach kritisch hinschauen und hinhören. Das geht jetzt vermehrt für digitale Medien. Wir sind gewohnt, das, was wir hören und sehen, für bare Münze zu nehmen. Wir müssen einfach verstehen, dass wir in der digitalen Welt nicht sofort dem Bild und dem Ton trauen können, ohne kritisch zu reflektieren. Wir sehen beispielsweise Werbeanzeigen, wo die Menschen geschönt sind, und das ist für uns vollkommen normal, weil man es einfach macht, damit sie gut rüberkommen. Solche Schönungseffekte gibt es genauso in anderen Bereichen, unter anderem im Audio- und Videobereich. Wir müssen lernen, Kritik aufzubringen, kritisch zu hinterfragen und in den Kontext stellen. Passt das Ganze in den Kontext? Das war, glaube ich, euer Einstieg in das Gespräch.

Münz: Gibt es darüber hinaus noch Tipps für Verbraucherinnen und Verbraucher, die du uns mit auf den Weg geben kannst? Gibt es vielleicht auffällige technische Aspekte?

Ullmann: Dafür muss man vielleicht in Richtung Schwachstellen oder Artefakte, die die Systeme verursachen, schauen. Wir bleiben jetzt beim Video. Es ist typischerweise so, dass in dem Video ein Gesicht ausgetauscht wird. Du, Michael, wurdest beispielsweise aufgezeichnet, und dann wird dein Gesicht durch meins ausgetauscht, obwohl unsere Gesichter und Haare ein bisschen unterschiedlich sind. Interessant sind immer die Randbereiche zwischen deinem und meinem Gesicht, die Übergänge zu den Haaren sowie wie es unten am Hals aussieht. Schwierigkeiten machen auch die Abbildungen der Zähne. Man müsste auf die Zähne schauen und merken, ob sie echt wirken oder ob man

Zahnstümpfe sieht. Danach guckt man auf die Synchronität von Mundbewegungen beim Sprechen. Können die zur Sprache passen? Sind sie synchron oder zeitverzögert? Das wären Punkte, die man Zuhörerinnen und Zuhörer mitgeben könnte, damit sie mit ihren Möglichkeiten gucken, ob es echt klingt und aussieht.

***Lange**:* Ich würde gerne noch mal auf dein Kaputtmachen und Wiederaufbauen vom Anfang zurückkommen. Es klang so, als ob euch das ziemlich viel Spaß macht. Ich sehe dich gerade wieder lächeln. Was war das schönste Kaputtmachen und Wiederaufbauen? Was habt ihr für IT-Sicherheit draus gelernt, von dem wir auch profitieren?

Ullmann: Jetzt muss ich ein bisschen nachdenken. Wenn wir diese Fakes machen, kommt immer das Problem vor, Personen zu finden, die bereit wären, mitzumachen. Am Ende hat unser Präsident das mit uns zusammen gemacht. Wir konnten seine Stimme aufnehmen, um die nachher synthetisch zu erzeugen. Das war interessant.

Münz: Falls ihr wieder einen Probanden oder eine Probandin braucht, sind Ute und ich bestimmt gerne dabei, würden gerne schauen, was da geht, und könnten in einer der späteren Folgen das vielleicht aufgreifen. Vielleicht finden wir auch irgendwann eine Möglichkeit, einen Podcast über euch synthetisch zu produzieren.

Ullmann: Herzlich gerne. Nehmen wir gerne an!

Münz: Danke dir! Für uns war hier viel drin. Ich glaube, wir können als Fazit festhalten, dass die Beachtung beider Bereiche wichtig ist, sowohl das Verständnis, was jemand versucht, mit mir zu tun oder mir zu vermitteln, als auch das Wissen über die technischen Hintergründe und Möglichkeiten. Das war für mich sehr deutlich an diesem Gespräch und ich werde das bei dem nächsten Mal, wenn ich mir eine Video- oder Tonaufnahme ansehe oder anhöre, berücksichtigen. Ute, und für dich? Was ist dein Takeaway heute?

Lange: Bei mir ist es wieder der Faktor Mensch. Wir haben es öfter angesprochen, und zwar wenn etwas zu unwahrscheinlich klingt, um wahr zu sein, ist es das wahrscheinlich auch. Wie du früher gesagt hast, Markus, man muss immer kritisch hinterfragen und mehrmals prüfen, ob beispielsweise meine Freundin wirklich jetzt auf dem Messenger Dienst ist, weil es mir komisch vorkommt. Ist das wirklich die Person, die mich anspricht und mir einen Megadeal oder Superschnäppchen anbietet? Man sollte vielleicht kurz durchatmen und nachdenken. Das habe ich mitgenommen. Vielen Dank, Markus. Ich fand das Gespräch sehr spannend, vor allem das, was heute schon möglich ist. Und wie du sagst, es ist nicht alles nur schlecht. Wir gewinnen eine ganze Menge Möglichkeiten in vielen Bereichen. Einige davon hast du uns vorgestellt. Herzlichen Dank! Damit sind wir schon bei der nächsten Folge. Michael, was haben wir für den nächsten Monat vorgesehen?

Münz: Für den nächsten Monat haben wir anlässlich der Gamescom das Thema Gaming vorgesehen. Wir schauen uns genauer an, wie Gaming und Sicherheit zusammengehen und wer sich vorab schlaumachen wollte. Wir haben im vergangenen Jahr dazu eine Folge aufgenommen, die sehr spannend war, und greifen das Thema jetzt im August noch einmal auf.

Lange: All die Informationen, die das BSI zum Thema Deepfakes und Sprachfakes hat, packen wir in die Shownotes für diejenigen, die das nachlesen und vielleicht auch den Spaß von Markus mit dem Präsidentenvideo nachvollziehen möchten. Bis wir uns wiederhören, liked und folgt Update Verfügbar auf euren Podcast Plattformen. So verpasst ihr keine Folge und könnt auch gerne ältere Folgen nachhören.

Münz: Wie immer gilt: Kontaktiert uns über die BSI-Kanäle auf Facebook, Instagram, Twitter und YouTube. Wir schauen da rein und greifen eure Rückmeldungen gerne auf. Oder schickt uns eine Mail an bsi@bsi.bund.de. Wir freuen uns auf Post!

Lange*:* Immer auf allen Kanälen bis wir uns wiederhören. Alles Gute, einen schönen Sommer, bleibt gesund und bis zur nächsten Folge. Tschüss!

Besuchen Sie uns auch auf:

https://www.bsi.bund.de/
https://www.facebook.com/bsi.fuer.buerger
https://twitter.com/BSI_Bund
https://www.instagram.com/bsi_bund/
https://social.bund.de/@bsi

Herausgeber: Bundesamt für Sicherheit in der Informationstechnik (BSI),
Godesberger Allee 185-189, 53133 Bonn