GUERRE EN UKRAINE ET DEEPFAKES : UN OUTIL DISCURSIF, STRATÉGIQUE ET SYSTÉMIQUE

Léo-Paul BARTHELEMY
leo-paul.barthelemy@univ-lorraine.fr
Université de Lorraine, France

Abstract: *The war in Ukraine has revealed the rise of digital visual manipulation with AI tools, marking a new era in informational warfare. Among them, deepfakes, sophisticated hypertrucages based on deep learning technologies, have emerged as powerful disinformation weapons. Broadcast on social networks or television, these manipulated contents exploit digital virality to maximize their reach. Propagandists from both sides leverage deepfakes to influence, deceive, manipulate public opinion and reinforce narratives. This exploratory article seeks to create an overview of deepfake trends in this war, from their creation to their dissemination, through representative examples illustrating their strategic role in digital warfare.*

The notable examples of deepfakes in our corpus offer insight into how both sides are using these tools and reveal several general observations. These synthetic images, which have limited diplomatic impact, nevertheless mark a small informational turning point in this context of hybrid warfare. Deepfakes are part of a systemic informational approach for both sides. Specifically, they support other campaigns, tools and techniques to influence the propaganda messages desired by both sides. The main commonalities include, among other things, demoralising the enemy, psychological attacks, emotional manipulation and the dissemination of deepfakes via strategic media such as television or popular social networks. In addition, the deepfakes analysed show that they are used to (de)mobilise populations, soldiers and sometimes even strategic allies, while seeking to (dis)credit the messages of both sides. The narratives used thus aim to psychologically influence target audiences, often leading them to believe that it is necessary to lay down their arms.

Keywords*: deepfakes, Russo-Ukrainian war, visual manipulation, infowar, AI.*

Introduction

L'avènement de la guerre hybride en Ukraine a fait ressurgir des aspects traditionnels de l'infoguerre, telles que la manipulation de l'information ou la guerre psychologique, déjà étudiés à travers les conflits précédents. Cependant, ce théâtre de guerre se distingue par sa matérialisation et son développement dans un espace numérique marqué par l'évolution de l'intelligence artificielle (IA). Qualifiée, peut-être hâtivement, de « First Tiktok War » (Kennedy, 2022), la guerre en Ukraine s'inscrit malgré tout dans une époque informationnelle. Parallèlement, les réseaux sociaux, comme Telegram, TikTok,

Instagram ou X, sont largement prisés pour relayer des informations, souvent visuelles, sur des sujets liés à ce conflit.

Avec l'essor des réseaux sociaux depuis les années 2010, les fausses informations prolifèrent rapidement. Parallèlement, la démocratisation des outils d'IA a permis au grand public de créer facilement et gratuitement des contenus artificiels, qu'ils soient textuels, visuels ou sonores. De ce contexte a pu émerger une pratique dite de « deepfake » : le remplacement de visages de personnes célèbres (médiatiques, politiques, cinématographiques) par un autre visage. Tout ceci repose sur les technologies d'apprentissage profond (« deep learning »), à l'aide des réseaux de neurones artificiels.

La guerre en Ukraine est souvent qualifiée de « guerre hybride » dans la plupart des discours politiques et médiatiques. Cette notion, popularisée par l'annexion de la Crimée en 2014 (Reichborn-Kjennerud & Cullen, 2016), se définit par une combinaison « des opérations de guerre conventionnelle, de guerre asymétrique (ou irrégulière), de cyberguerre et d'autres outils tels que la désinformation » (Desportes et al., 2023). Les deepfakes (hypertrucages) occupent une place de plus en plus importante dans notre quotidien numérique et informationnel. Néologisme né de la contraction de deep learning (apprentissage profond) et fake (faux), les deepfakes sont « des contenus synthétiques créés avec une intelligence artificielle afin de produire des contenus synthétiques réalistes » (Dugoin-Clément, 2020) qui seraient un « équivalent visuel des infox » (Allard, 2021). Denis Teyssou, s'appuyant sur les travaux de l'influent universitaire Umberto Eco, rappelle que ces images synthétiques « n'ont ni origine traçable ni réalité » (Teyssou, 2024).

La littérature scientifique sur les deepfakes, bien que récente, existe mais se consacre plutôt à ses utilisations dans des contextes classiques. De nombreux universitaires et philosophes ont exprimé une crainte quant aux menaces posées par les deepfakes sur la véracité des informations diffusées (Fallis, 2021), bien que ces commentaires soient souvent hypothétiques, du fait de l'absence de cas concret de dommages politiques et épistémologiques causés par ceux-ci (Twomey et al., 2023). Notons que la notion de deepfakes fait l'objet de définitions plurielles, avec tantôt une signification limitée au « swap » de visages de célébrités, tantôt une notion plus englobante tous types de contenus générés à l'aide de l'intelligence artificielle. À ce jour, une seule étude sur l'utilisation de deepfakes dans le cadre de la guerre en Ukraine existe. Menée par des chercheurs de l'University College Cork of Ireland (Twomey et al., 2023), leur enquête a mis en avant un constat pour le moins inattendu : les contenus proposés aux enquêtés étaient souvent interprétés comme faux alors qu'ils ne l'étaient pas. Des vidéos réelles étaient prises pour des deepfakes. En outre, les chercheurs ont révélé que ceci se traduisait par une méfiance croissante envers les médias ainsi qu'un renforcement des théories complotistes, les internautes ne sachant plus discerner le vrai du faux dans le brouillard de guerre numérique.

Au croisement de l'avènement des outils recourant à l'intelligence artificielle et d'un contexte de guerre hybride en Ukraine où l'espace numérique occupe une place de premier plan, quelle place occupent les deepfakes dans les stratégies informationnelles des deux camps ? Ces usages présentent-ils des tendances communes ou divergentes, et le cas échéant, lesquelles ?

L'objectif de cette contribution est de comparer les stratégies discursives des belligérants ukrainiens et russes à travers les deepfakes, dans le but de dégager les intentions et objectifs de chacun. Cette analyse réflexive repose sur une approche descriptive et comparative, en étudiant les deepfakes des deux camps, médiatisés entre 2022 et 2024 et ayant eu une forte couverture dans la presse internationale. Nous nous

concentrerons sur des éléments variés, tels que les visuels, les tonalités, ou encore les intentions principales des créateurs, en fonction des caractéristiques propres à chaque deepfake, afin d'identifier les tendances manipulatoires inscrites dans les stratégies discursives des deux camps. Nous tenterons également de comprendre si ces deepfakes s'inscrivent dans des temporalités particulières ou si leur diffusion suit une logique plus aléatoire. Enfin, nous chercherons à mettre en évidence des tendances susceptibles de constituer des pistes pour des recherches plus approfondies à l'avenir.

En dépit des préoccupations croissantes, la médiatisation de deepfakes dans ce contexte de guerre reste marginale, avec peu de cas concrets comparés à l'agitation médiatique qu'ils suscitent. S'ils font l'objet d'inquiétudes et d'usages variés par les deux camps, leur cycle de vie semble suivre une trame récurrente que nous déduisons à partir des exemples de notre corpus.

1. Création du deepfake avec l'utilisation d'outils de deep learning.
2. Amélioration éventuelle à l'aide d'outils d'édition et/ou d'aide humaine.
3. Diffusion sur les réseaux sociaux, à la télévision, ou dans la presse en ligne.
4. Amplification de la diffusion en recourant au botting, au spamming, au trolling et en jouant avec la puissance des algorithmes.
5. Réception par les publics cibles : impact émotionnel et/ou psychologique.
6. Conséquences variables : du raté à la méfiance vis-à-vis de l'émetteur ou du relai.

Méthodologie

Pour cette recherche, nous recourons à un corpus de deepfakes créés et diffusés par des acteurs – souvent non identifiés – pro-russes et pro-ukrainiens. À l'aide de cette analyse comparative et réflexive, nous mettrons en évidence les objectifs et enjeux stratégiques souhaités des deux camps. Nous avons sélectionné huit exemples (quatre pour chaque belligérant), numérotés et dont les liens de visionnages sont disponibles à l'issue de l'article, en nous basant sur leur résonance médiatique. Pour ce faire, la recherche dans les moteurs de recherche avec les termes génériques « deepfakes », « guerre », « Ukraine », « Russie » (ainsi que leurs équivalents en anglais et en ukrainien) a permis de faire ressortir les articles de presse numériques qui évoquaient les cas les plus médiatisés et discutés jusqu'à présent. Bien que des deepfakes soient créés quotidiennement, la majorité d'entre eux ont une durée de vie courte et tombent rapidement dans l'oubli, ne dépassant souvent pas la quatrième étape du cycle des deepfakes évoqué précédemment. Nous en évoquerons un exemple dans les discussions de cet article. Le nombre de deepfakes ayant eu une résonance médiatique reste faible, malgré presque trois années écoulées depuis l'invasion à grande échelle. Nous étudierons les deepfakes, d'abord du camp russe puis ukrainien, selon plusieurs critères : le contexte de création et de diffusion, les cibles visées, les narratifs dominants véhiculés, les objectifs stratégiques poursuivis ou encore les réactions observées, lorsqu'il y en a. Ces résultats préliminaires, synthétisés dans un tableau, offrent un aperçu des narratifs dominants des deux camps, ainsi que des similitudes et différences du corpus.

Diviser pour mieux régner : les deepfakes au service de la Russie

La stratégie propagandiste russe jouit d'une littérature scientifique substantielle, et ce, depuis l'époque soviétique. Les deepfakes créés par les Russes s'inscrivent dans une stratégie informationnelle élargie et systémique, où règnent le multicanal et les gros volumes d'informations (Paul & Matthews, 2016). Dès 2022, l'opération de désinformation « Doppelgänger » est révélée, avec près de 50 faux sites usurpant l'identité graphique de

médias européens. Plus récemment, en 2023, la seconde phase de cette campagne de désinformation s'est poursuivie sous le nom de « Portal Kombat ». Il s'agissait ici de légitimer les narratifs russes en relayant en abondance des narratifs inexacts.

[1] Le premier deepfake russe à avoir été massivement médiatisé est celui du discours du président Zelensky, le 16 mars 2022, suite à un piratage de la chaîne de télévision et du site web d'Ukraine 24. À peine un mois après le début de l'invasion à grande échelle opérée par la Russie, ce premier cas de deepfake marquait une pierre angulaire dans la guerre d'information. Pendant près d'une minute, le président ukrainien appelle ses concitoyens à rendre les armes. La vidéo, relayée d'abord sur VKontakte et Telegram, puis sur Facebook, Instagram et X, a été diffusée alors que les troupes russes approchaient de Kyiv, visant à renforcer un discours défaitiste. Le discours est de toute évidence conçu pour atteindre la population ukrainienne : le président se tient l'air grave, face à un pupitre, et évoque en de termes simples et directs la nécessité d'arrêter les combats. Il recommande de « déposer les armes » et de « retourner auprès de vos familles », tout en ajoutant « je vous demande de vivre, je compte faire de même » (notre trad.). Ses injonctions sont directes et le discours est prononcé de manière à ce que le peuple suive son exemple. L'effet escompté ne sera évidemment pas au rendez-vous du fait de la piètre qualité du deepfake. Malgré la large diffusion de ce deepfake, ses résultats ont été limités en raison de sa mauvaise qualité, notamment l'accent et la synchronisation labiale défaillants. Le démenti rapide du président ukrainien Zelensky *via* une vidéo selfie sur Telegram a également contribué à désamorcer son impact. Ce deepfake, le premier massivement diffusé dans un conflit moderne, a attiré l'attention des médias internationaux, soulignant la vigilance nécessaire face à ce type de manipulation visuelle.

[2] Le deuxième deepfake de notre corpus concerne un faux clip publié sur les réseaux sociaux, principalement TikTok, X et Telegram, montrant la ville de Paris assiégée et en flammes suite à des bombardements. Créée et diffusée par les Ukrainiens pour leur propre agenda informationnel, elle a été réutilisée par les Russes pour retourner l'argumentaire. La vidéo a été créée expressément pour faire réagir les Occidentaux (dans le but de les sensibiliser à la guerre), avant d'être reprise par la suite par des relais russes qui se sont appuyés sur son caractère fictif pour pointer les dangers de l'intelligence artificielle. Cette réutilisation discursive montre qu'un même deepfake peut être employé par deux adversaires en jouant avec les interprétations et les intentions voulues. La lecture est donc double pour les internautes : les Occidentaux peuvent se sentir davantage concernés par la guerre avec ce narratif de proximité proposé par les Ukrainiens, tout en se sentant davantage troublés par les arguments appuyés par les Russes quant aux dangers des manipulations de contenus. De ce fait, bien que ce deepfake soit à l'initiative des Ukrainiens, nous comprenons qu'il peut s'avérer tout à fait utile pour alimenter les narratifs russes.

[3] Notre troisième exemple propulsé par des pro-russes est celui de combattants ukrainiens de la 117e brigade de défense territoriale, qui annoncent se rendre. Ce faux, diffusé sur TikTok en février 2024, montre une nouvelle fois que son inscription temporelle n'est pas le fruit du hasard. La vidéo a été diffusée massivement au moment où des discussions étaient en cours sur les conditions de la mobilisation des conscrits en Ukraine. Les propagandistes, à la tête de plusieurs canaux Telegram pro-russes (Panchenko, 2024), ont cherché à marquer le coup en créant ce deepfake. Les faux visages sont ajoutés de manière grossière sur les soldats et les commentaires des internautes suggèrent que le deepfake ne convainc pas beaucoup. Ceci étant, à l'instar du premier exemple du corpus, le narratif voulu est bien celui de la démobilisation. Le faux a provoqué une réaction par la

brigade elle-même, qui s'est justifiée sur Facebook en démentant, arguments à la clé, la supercherie. Le démenti conclu avec « continuons à croire en nos Forces de Défenses » (notre trad.), montrant que le deepfake est perçu comme une menace pour l'unité et le soutien moral aux forces ukrainiennes.

[4] Le quatrième exemple, différent des précédents de par sa nature et sa portée, est tout aussi révélateur des enjeux posés par les deepfakes dans ce contexte d'infoguerre international. Pendant plusieurs mois en 2023, un créateur de contenu sur l'équivalent chinois de TikTok (Douyin) s'est fait passer pour un soldat tchétchène en mission en Ukraine. L'auteur du compte utilisait le pseudonyme Baoer Kechatie ; nom inspiré de Pavel Kortchaguine, personnage central du célèbre roman soviétique *Et l'acier fût trempé* de Nikolaï Ostrovski. Il prétendait devant sa communauté de plus de 400 000 abonnés qu'il avait remporté des combats contre des Marines américains ou qu'il avait participé à la capture de la centrale nucléaire de Zaporijia. Tout ceci s'est avéré être faux, son accent chinois étant rapidement identifié et son adresse IP localisée en Chine. Clint Watts, directeur général du Microsoft Threat Analysis Center, justifie cet échec en disant que les Chinois « ont du mal avec le contexte culturel » (notre trad.) (Sydney J. Freedberg Jr, 2024). Du fait de sa popularité, il aura tout de même réussi à réaliser des ventes e-commerce de produits russes, comme du miel ou de la vodka. Ses faux, visiblement motivés par une initiative personnelle, semblent déconnectés des intérêts des autorités russes. Cela pose la question des dérives des deepfakes en dehors des stratégies informationnelles censées être maîtrisées. Ce cas montre que l'usage des deepfakes en temps de guerre dépasse le champ de bataille, permettant à quiconque d'en créer, même à des milliers de kilomètres du conflit. Les propos de ces vidéos, indirectement adressés aux Américains, illustrent une utilisation alternative des deepfakes : un émetteur chinois, se faisant passer pour un soldat russe en Ukraine, exprime sa rancœur à l'encontre des États-Unis. Cette approche dépasse l'usage habituel, où les deepfakes ciblent directement les acteurs du conflit (Ukrainiens et Russes), pour s'adresser à un public international, y compris chinois, et élargir l'impact de la désinformation.

De la mobilisation à la dénonciation : la stratégie pro-ukrainienne

[5] Le président russe, Vladimir Poutine, n'a pas non plus été épargné par les deepfakes. Plusieurs faux ont été réalisés par les partisans pro-ukrainiens. Le premier cas notable est celui où il annonçait la fin de la guerre dès le mois de mars 2022. La vidéo, relayée sur X et Telegram par l'influent activiste ukrainien Serhii Sternenko, se voulait être une réponse à celle diffusée peu de temps avant sur Zelensky. D'une qualité un peu supérieure, quelques décalages labiaux viennent toutefois confirmer le recours à l'intelligence artificielle. En russe, le président parle de « paix »» et n'évoque toutefois pas de capitulation comme l'annonce le titre de la vidéo. Plusieurs points sont contradictoires et viennent crédibiliser le montage : la vidéo annonçait la paix au moment où le conflit gagnait en intensité et il n'y avait pas d'écho médiatique évoquant cette paix. De plus, si le titre du message de Sternenko semble s'adresser aux soldats, le discours quant à lui est plus vaste et s'adresse aux Russes dans leur globalité, avec un ton qui se veut rassurant. Les images utilisées n'étaient ni plus ni moins que celles du discours de Vladimir Poutine du 21 février 2022 lorsqu'il reconnaissait l'indépendance des régions de Lougansk et Donetsk. L'activiste commentera même, sous sa propre vidéo sur le réseau social X : « Apprenez à faire des deepfakes, katsaps. C'est un travail de qualité, pas les conneries que vous avez inventées contre Zelensky. » (notre trad.). Cette dynamique de « deepfake-réponse » sur les

présidents n'a pas duré, les deux camps se concentrant ensuite sur des événements liés à leurs efforts de guerre respectifs.

[6] À ce propos, Vladimir Poutine est de nouveau ciblé par un deuxième deepfake relayé massivement le 5 juin 2023. Le timing n'est pas laissé au hasard puisqu'il s'agit du moment où la contre-offensive ukrainienne a eu lieu. Cette fois-ci, le discours est partagé plus rapidement sur les réseaux sociaux et par plusieurs médias, y compris en Russie. La réponse ne se fait pas attendre côté russe : les autorités estiment qu'il s'agit de « semer la panique » et Dmitri Peskov tempère en disant « qu'il n'y a pas eu d'adresse d'urgence à la télévision » tout en évoquant « un piratage ». Cette fois-ci, le discours de Vladimir Poutine annonce l'évacuation de trois régions (Koursk, Belgorod, Briansk). Il ajoute : « je signerai aujourd'hui un décret pour la mobilisation générale ». En ayant recours à la première personne et en utilisant des termes précis, les créateurs du deepfake souhaitent de toute évidence troubler la population locale pour les déstabiliser. On y entend le président dire « nous devons unir nos forces pour vaincre cet ennemi ». Ce deepfake s'inscrit à nouveau à un moment crucial de la guerre, quelques jours après les premières incursions dans la région de Belgorod, initiées dès le 22 mai 2023. Cet exemple nous montre bien à nouveau que ces réalisations se calquent sur la réalité des événements en cours, dans une logique propagandiste plus vaste.

[7] Le troisième exemple de deepfake du président russe est celui diffusé pour son 72e anniversaire, le 7 octobre 2024, sur la chaîne de Crimée Krym24 TV. Dans un premier temps, un groupe de pirates informatiques pro-Ukraine, Sudo rm -RF, a massivement attaqué l'un des plus gros mastodontes russes médiatiques, VGTRK. Cela en a résulté en une interruption des programmes, la destruction de nombreuses archives du groupe ou encore le dysfonctionnement de serveurs. S'attaquer à ce maillon principal du récit russe répond à un double objectif : symbolique, d'une part, puisque cela montre la vulnérabilité dans le cyberespace de cette entité, stratégique, d'autre part, puisque cela permet de toucher efficacement un grand nombre de personnes simultanément. Dans un second temps et pendant presque une minute, un deepfake s'adressant à la population de Crimée et des « nouveaux territoires » a été diffusé. Poutine annonce qu'il s'agit de son dernier anniversaire comme président, et que les Ukrainiens « délivrent des frappes précises sur les troupes russes » (notre trad.). Et d'ajouter : « le gouvernement ne peut plus assurer ni la sécurité ni la stabilité » (notre trad.). L'intention des hackers pour la population est claire. Sur un ton impératif et éminemment calme, le président russe « appelle » les habitants « à ne pas résister ». Le message se conclut même par un appel à la collaboration en interpellant les auditeurs : « Vous pouvez faciliter ce processus en fournissant des informations cruciales aux forces armées ukrainiennes pour arrêter les combats. » (notre trad). Outre la multiplicité des objectifs, cet exemple rappelle bien l'ancrage systémique des deepfakes dans des techniques d'infoguerre plus vaste. Ici, il s'agit d'un recours à des techniques de cyberattaques. Le deepfake n'est finalement qu'une finalité dans un maillon informationnel bien plus étendu dans cette opération. Le porte-parole russe Dmitri Peskov a réagi, montrant la gravité de la situation et la prise au sérieux de celle-ci par les autorités russes.

[8] Il est intéressant de voir la réaction du principal concerné : là où Volodymyr Zelensky avait démenti à travers une vidéo selfie, Vladimir Poutine a été confronté à son propre deepfake dans le cadre de sa conférence télévisée annuelle en décembre 2023. Un étudiant lui a demandé : « Est-ce que vous avez beaucoup de doubles ? » (notre trad.). À ceci, le – vrai – président russe a réagi : « Nous ne pouvons pas l'éviter. » (notre trad.). Il

ajoute d'un ton calme : « Je vois que vous me ressemblez et que vous parlez avec ma voix, mais j'ai réfléchi, et je pense qu'il ne doit y avoir qu'une personne qui me ressemble et qui ai ma voix. Et cette personne, c'est moi ! » (notre trad.). L'objectif de sa réponse – qu'elle soit préparée ou non – est certainement multiple : démontrer sa compréhension des enjeux de l'intelligence artificielle, alerter indirectement la population sur l'existence de tels deepfakes, présenter une image maîtrisée en recourant à l'humour ou encore couper court aux anciennes rumeurs qui imaginaient que le président russe est remplacé par un clone. Enfin, son discours met l'accent à la fois sur le physique et la voix, soit les deux éléments clés d'un deepfake, que le président russe cherche à souligner pour avertir l'audimat.

Comparaison des tendances et stratégies générales observées

La comparaison de ces exemples de deepfakes met en exergue des stratégies relativement distinctes, bien qu'elles partagent quelques points communs. Spécifiquement, les objets visés (politiciens et soldats) et les narratifs principaux (mobilisation et démobilisation) employés de part et d'autre se rejoignent dans les objectifs poursuivis, malgré les différences de contexte et de cible.

La stratégie discursive côté russe repose sur une posture d'abondance, visant à diviser l'opinion publique internationale quant au conflit. Elle s'ancre dans une longue tradition de propagande, forte d'un maillage numérique mondial, appuyée par des campagnes de désinformation massives pour ses messages. Selon Clint Watts, les propagandistes russes ont rapidement abandonné les deepfakes pour retourner aux manipulations visuelles classiques (Freedberg Jr, 2024). Ceci est peut-être une première explication sur leur faible proportion dans l'espace médiatique. Ces deepfakes sont principalement orientés vers des cibles multiples : les soldats et la population ukrainienne, mais aussi les alliés de l'Ukraine par extension. En ce sens, le cas où le président russe « discute » avec son « double » montre une maîtrise des enjeux de l'IA. Les autorités russes savent qu'un tel événement sera relayé et commenté à l'étranger. Ces deepfakes sont diffusés principalement pour semer la confusion, affaiblir la diplomatie ukrainienne et démoraliser l'adversaire. En réponse aux deepfakes propagés par les propagandistes opposés, la Russie dément souvent ces faux *via* ses porte-paroles, cherchant à protéger la légitimité des autorités.

De son côté, la stratégie ukrainienne adopte une posture défensive et réactive, axée sur la contre-attaque symbolique et la sensibilisation de ses alliés. L'Ukraine mise accessoirement sur des narratifs identitaires et patriotiques, mais l'objectif premier reste de démobiliser l'adversaire et de décrédibiliser les autorités russes. De ce fait, le président Vladimir Poutine est très souvent au centre de ces manipulations par IA. Les canaux privilégiés pour la diffusion sont similaires à ceux de la Russie, avec une forte présence sur Telegram, X, Instagram et TikTok, et dans une moindre mesure VKontakte et Rutube, très prisés en Russie. Les réactions ukrainiennes aux deepfakes russes incluent des efforts de « dé-bunkage », avec des démentis diffusés sur les réseaux sociaux pour contrer les fausses informations, à l'instar de la 117e brigade motorisée comme nous l'avons vu.

Les objectifs des deux camps sont, dans les grandes lignes, similaires : manipuler l'opinion publique du pays adversaire, décrédibiliser la parole présidentielle ou faire ressurgir un sentiment d'abandon en inventant de fausses redditions. Les deepfakes s'intègrent à des stratégies élargies, combinant cyberattaques, spamming de fake news et piratages de chaînes télévisées, ces dernières servant de vecteurs symboliques. Néanmoins, leurs approches diffèrent sensiblement. Le camp russe privilégie une stratégie de masse et de division, tandis que le camp ukrainien se concentre sur des actions davantage symboliques et ciblées.

Décorréler les deepfakes des événements de la guerre serait potentiellement inefficace pour jouir d'une couverture médiatique.

Objets des deepfakes	Événements clés associés	Cibles principales	Narratifs principaux	Intention discursive
[1] *Discours de Zelensky* (03/2022)	Troupes russes aux abords de Kyiv	Soldats et pop. ukrainiens	Démobilisation, capitulation	Décrédibiliser
[2] *Ville de Paris assiégée* (03/2022)	Appel à l'aide internationale	Alliés de l'Ukraine	Désespoir, mobilisation	Crédibiliser
[3] *117e brigade mécanisée* (02/2024)	Discussions sur la mobilisation en Ukraine	Soldats et pop. ukrainiens	Démobilisation, capitulation	Décrédibiliser
[4] *Influenceur chinois se faisant passer pour un soldat russe* (2023)	Internationalisation de la guerre hybride	Internationale (États-Unis, Chine)	Déstabilisation, influence, mobilisation	Crédibiliser
[5] *Discours de V. Poutine* (03/2022)	Réponse au deepfake de V. Zelensky	Pop. Russe	Démobilisation, capitulation	Décrédibiliser
[6] *Discours de V. Poutine* (06/2023)	Contre-offensive ukrainienne 2023	Pop. Russe	Démobilisation, capitulation	Décrédibiliser
[7] *Discours de V. Poutine* (10/2024)	Anniversaire de V. Poutine	Pop. Russe	Démobilisation, capitulation	Décrédibiliser
[8] *Discours de V. Poutine* (12/2023)	Grande conférence annuelle de V. Poutine	Pop. Russe, Internationale	Pouvoir, contrôle, influence	Crédibiliser

Quatre tendances générales ressortent au regard des exemples de notre corpus :

1. Les deepfakes s'inscrivent systématiquement dans des moments clés du conflit.
2. Les responsables politiques en sont, très souvent, les objets principaux.
3. Les narratifs principaux tournent autour de la mobilisation (des siens) ou la démobilisation (de l'ennemi).
4. Chaque deepfake sert à crédibiliser ou à décrédibiliser les discours, selon l'émetteur.

Discussions

Comme l'ont souligné les chercheurs irlandais (Linehan et al., 2023), ces hypertrucages ne sont pour l'heure pas spécialement répandus en ligne. De plus, le journaliste pour BBC Monitoring, Shayan Sardarizadeh, rappelle que les anciennes vidéos et les mèmes falsifiés restent la tactique la plus courante dans cette guerre (Wakefield & Sardarizadeh, 2022). De ce fait, il est préférable d'appréhender le deepfake comme une évolution et non une révolution (Whyte, 2020). Par ailleurs, plusieurs limites sont à prendre en compte pour l'appréhension de cette étude.

Premièrement, ce corpus permet de dégager des tendances générales, mais il pourrait être enrichi par l'inclusion des « cheap fakes », des deepfakes créés dans un but humoristique et facilement perçus comme tels.

Deuxièmement, bien qu'ils ne soient pas intégrés à notre corpus en raison de leur diffusion plus limitée par la presse internationale, soulignons que d'autres responsables politiques ukrainiens et russes, comme Valeri Zaloujny, Vitali Klitschko et Maria Zakharova, ont également été visés par des deepfakes. Si ces contenus ne sont pas directement accessibles, plusieurs sources ouvertes font état de leur existence.

Troisièmement, les régulations vis-à-vis du réseau social Telegram, omniprésent dans le paysage informationnel de la guerre en Ukraine, nous contraint à ne plus accéder à

certains canaux, notamment russes, tels que ceux de Russia Today (RT). De ce fait, il est plus difficile de suivre le cycle de vie des deepfakes dans l'espace numérique russe.

Par ailleurs et de façon contre-intuitive, pouvons-nous supposer que plus un deepfake est méconnu, plus il pourrait s'avérer dangereux ? Lorsque sa diffusion est restreinte et que son cycle de vie est « incomplet », les quelques personnes exposées pourraient avoir davantage de mal à démentir l'information, en particulier si elles ne disposent ni des codes ni des outils nécessaires. À titre d'exemple, on estime qu'il y aurait eu une trentaine de deepfakes diffusés lors du premier mois de l'invasion ukrainienne dans la région de Koursk (Solovian & Wickham, 2024). Peu d'informations circulant sur ces créations, il n'est pas possible de savoir comment la population locale, de surcroît rurale et potentiellement plus vulnérable, a pu percevoir ces hypertrucages. Cela nous interroge plus que jamais sur la nécessité du *flair sémiologique*, cette capacité à saisir du sens là où l'on se contenterait de simples faits, comme le proposait Roland Barthes. Ceci, en plus de l'exemple chinois susmentionné, suggère que les propagandistes ont tout intérêt à prendre en compte les contextes sociaux et culturels pour maximiser l'efficacité de leurs deepfakes.

Enfin, notons aussi que les réseaux sociaux constituent le terreau idéal pour la prolifération des deepfakes : de l'anonymisation des messages aux fonctionnalités de partage en passant par les algorithmes puissants des plateformes, tout est présent pour une diffusion rapide en ligne, notamment à une époque où 70 % des citoyens utilisent Telegram pour s'informer sur la guerre (Vorotyntseva, 2024).

Conclusions

Les exemples notoires de deepfakes de notre corpus offrent un aperçu des tendances d'utilisation de ces outils par les deux belligérants et révèlent plusieurs constats généraux. Ces images synthétiques, à l'impact diplomatique limité, marquent malgré tout un petit tournant informationnel dans ce contexte de guerre hybride. Les deepfakes s'inscrivent dans une démarche informationnelle systémique pour les deux camps. Précisément, ils viennent en appui à d'autres campagnes, outils et techniques pour influencer les messages propagandistes souhaités de part et d'autre. La dissémination des deepfakes est difficilement traçable au début, il est plus aisé d'attribuer un deepfake à un « camp » qu'à ses créateurs.

Les points communs majeurs incluent, entre autres, la démoralisation de l'ennemi, l'attaque psychologique, la manipulation émotionnelle et la diffusion des deepfakes *via* des médias stratégiques comme la télévision ou des réseaux sociaux très populaires. En outre, les deepfakes analysés montrent qu'ils servent à (dé)mobiliser les populations, les soldats et parfois même les alliés stratégiques, tout en cherchant à (dé)crédibiliser les messages des deux camps. Les narratifs utilisés visent ainsi à influencer psychologiquement les publics cibles, en leur faisant souvent croire qu'il est nécessaire de rendre les armes.

Toutefois, certaines spécificités singulières et secondaires ressortent : les Russes ciblent davantage les audiences internationales de façon indirecte, cherchant à affaiblir les alliances diplomatiques de l'Ukraine, tandis que les Ukrainiens tentent plutôt de concentrer leurs efforts en tentant de saboter la légitimité de l'autorité présidentielle russe. Dans les deux cas, pour ce corpus, nous découvrons que les deepfakes ayant le plus d'écho médiatique sont systématiquement transposables à des événements majeurs et ne sont pas diffusés par hasard.

En sus, il serait pertinent d'étudier comment ces contenus sont perçus par les publics selon la typologie de chaque réseau social. Il serait également judicieux de mettre en

perspective les deepfakes avec d'autres contenus générés par l'IA, tels que les faux textes. La démocratisation d'outils recourant à l'intelligence artificielle amènera inévitablement d'autres réflexions au fur et à mesure qu'ils gagneront en sophistication, y compris dans un terrain informationnel aussi sensible que celui de la guerre.

REFERENCES

ALLARD, Laurence, (2021), *L'art peut-il résister aux robots et à l'intelligence artificielle ?*, dans NECTART 12, nº 1, pp. 146-153, disponible en ligne : https://doi.org/10.3917/nect.012.0146.

DESPORTES, Vincent ; DUPUY, Emmanuel, et BOISSELIER, Patrick, (2023), *Guerre hybride et risques transverses*, dans « Revue Défense Nationale », nº HS12, pp. 49-56, disponible en ligne : https://www.defnat.com/e-RDN/vue-article-cahier.php?carticle=621&cidcahier=1321.

DUGOIN-CLÉMENT, Christine, (2020), *Les "deepfakes", ces fausses vidéos créées pour nous influencer*, dans « The Conversation », disponible en ligne : http://theconversation.com/les-deepfakes-ces-fausses-videos-creees-pour-nous-influencer-131783.

FALLIS, Don, (2021), "The Epistemic Threat of Deepfakes", dans "Philosophy & Technology", 34, nº 4, pp. 623-43, disponible en ligne : https://doi.org/10.1007/s13347-020-00419-2.

KENNEDY, Liam, (2022), "Is Russia's invasion of Ukraine the first TikTok war?", dans "The Irish Times", disponible en ligne : https://www.irishtimes.com/opinion/is-russia-s-invasion-of-ukraine-the-first-tiktok-war-1.4884044.

LINEHAN, Conor ; MURPHY, Gillian, et TWOMEY, John Joseph, (2023), "Deepfakes in warfare: new concerns emerge from their use around the Russian invasion of Ukraine", dans "The Conversation", disponible en ligne : http://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393.

PANCHENKO, Ihor, (2024), "*Focha — is a deepfake*. A network of channels spreading AI videos *of Ukrainians* from the frontline is exposed on Telegram", dans ITC.ua (blog), disponible en ligne : https://itc.ua/en/news/focha-is-a-deepfake-a-network-of-channels-spreading-ai-videos-of-ukrainians-from-the-frontline-is-exposed-on-telegram/.

PAUL, Christopher, et MATTHEWS, Miriam, (2016), "The Russian "Firehose of Falsehood" Propaganda Model: Why It Might Work and Options to Counter It", dans "RAND", disponible en ligne : https://www.rand.org/pubs/perspectives/PE198.html.

REICHBORN-KJENNERUD, Erik, et CULLEN, Patrick, (2016), "What is Hybrid Warfare?", dans "Policy Brief", Norwegian Institute of International Affairs (NUPI), No 1, disponible en ligne : https://www.jstor.org/stable/resrep07978.

SOLOVIAN, Volodymyr, et WICKHAM, Matt, (2024), "The New Face of Deception: AI's Role in the Kremlin's Information Warfare", dans "UA: Ukraine Analytica", 3, nº 35, pp. 4-5, disponible en ligne : https://ukraine-analytica.org/wp-content/uploads/Solovian2.pdf.

FREEDBERG JR., Sydney J., (2024), "Deepfakes Deepfail: Russian Propagandists Turn Away from Generative AI", dans "Breaking Defense", disponible en ligne : https://breakingdefense.com/2024/09/deepfakes-deepfail-russian-propagandists-turn-away-from-generative-ai/.

TEYSSOU, Denis, (2024), « La vérification de l'information visuelle au défi de l'intelligence artificielle générative », dans « I2D - Information, données & documents », 242, nº 2, pp. 82-86, disponible en ligne : https://doi.org/10.3917/i2d.242.0082.

TWOMEY, John, CHING, Didier, AYLETT, Matthew Peter, QUAYLE, Michael, LINEHAN, Conor, et MURPHY, Gillian, (2023), "Do Deepfake Videos Undermine Our Epistemic Trust? A Thematic Analysis of Tweets That Discuss Deepfakes in the Russian Invasion of Ukraine", dans PLOS ONE 18, nº 10, disponible en ligne : https://doi.org/10.1371/journal.pone.0291668.

VOROTYNTSEVA, Maryna, (2024), "Russia's War in Ukraine: Russia's Attempts to Undermine Mobilisation", dans ICDS, disponible en ligne : https://icds.ee/en/russias-attempts-to-undermine-mobilisation/.

WAKEFIELD, Jane, et SARDARIZADEH, Shayan, (2022), "Deepfake Presidents Used in Russia-Ukraine War", dans BBC, disponible en ligne : https://www.bbc.com/news/technology-60780142.

WHYTE, Christopher, (2020), "Deepfake News: AI-Enabled Disinformation as a Multi-Level Public Policy Challenge", dans "Journal of Cyber Policy", volume 5, pp. 199-217, disponible en ligne : https://www.tandfonline.com/doi/abs/10.1080/23738871.2020.1797135.

Corpus de travail :

[1] Youtube – The Telegraph https://youtu.be/X17yrEV5sl4?si=zdqEJKws8WZbXF40 (consulté le 22/01/2025)
[2] Twitter – Verkhovna Rada of Ukraine – Ukrainian Parliament https://x.com/ua_parliament/status/1502402021386858504 (consulté le 12/01/2025)
[3] TikTok – ukrglass, @ukrglass https://www.tiktok.com/@ukrglass/video/7335361313985547538 (consulté le 12/01/2025)
[4] Rutube – RT https://rutube.ru/video/29e60cba855f9e3d494de7c0dfc27e52/ (consulté le 22/01/2025)
[5] Twitter – Serhii Sternenko, @sternenko https://x.com/sternenko/status/1504092439665164290 (consulté le 14/01/2025)
[6] Twitter – Hanna Liubakova, @HannaLiubakova https://x.com/HannaLiubakova/status/1665690283776802818 (consulté le 22/01/2025)
[7] Twitter – KyivPost, @KyivPost https://x.com/KyivPost/status/1843333219632161276 (consulté le 18/01/2025)
[8] Youtube – Guardian News https://youtu.be/KbaKTz9FW2E?si=TwDDYXaVZjzwExEI (consulté le 16/01/2025)