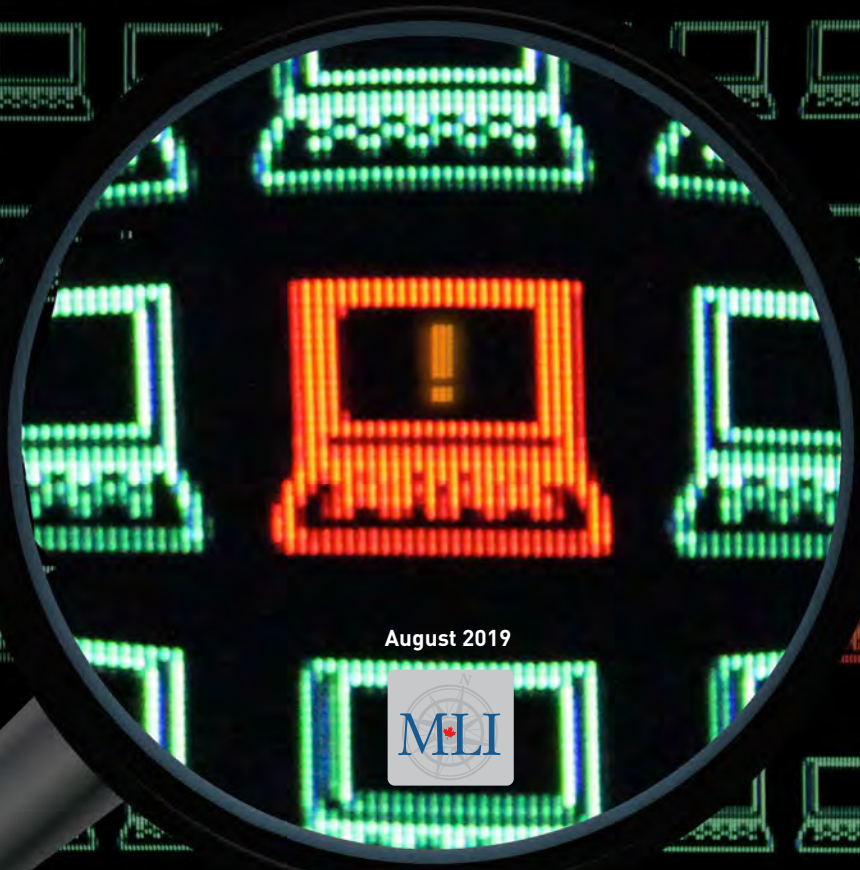
A Macdonald-Laurier Institute Publication

THE THREAT OF DIGITAL FOREIGN INTERFERENCE

PAST, PRESENT AND FUTURE

Alex Wilner
James Balasch, Jonathan Kandelshein,
Cristian Lorenzoni, Sydney Reis

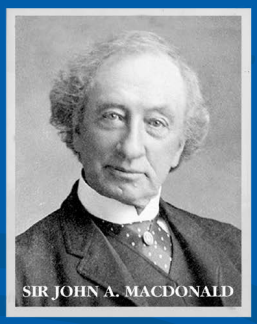


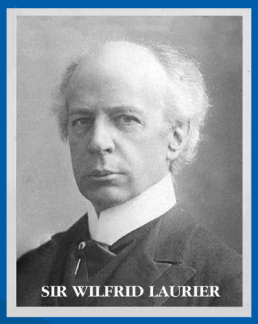
August 2019

MLI

MLI

True North in Canadian public policy


SIR JOHN A. MACDONALD


SIR WILFRID LAURIER

Board of Directors

Advisory Council

Research Advisory Board

Table of contents

The authors of this document have worked independently and are solely responsible for the views presented here. The opinions are not necessarily those of the Macdonald-Laurier Institute, its Directors or Supporters.

Executive Summary

Digital foreign interference (DFI) is an emerging Canadian national security challenge that is rising in significance as more malicious foreign actors learn how to effectively employ it against democracies. DFI is particularly dangerous because foreign actors can surreptitiously undercut established institutions, values, and norms by directly connecting with citizens in other countries through free and widely accessible methods of communication.

DFI is an evolved form of traditional propaganda, whereby the Internet and other types of technology are utilized to create and proliferate particular information in an immediate, targeted, and tailored way. As opposed to traditional forms of propaganda, DFI is employed primarily through the Internet and specifically via social media platforms. Once a malicious actor is virtually connected with foreign individuals and communities, they can create and disseminate tailored and targeted propaganda. In some cases, this propaganda feeds the national objectives of a foreign government, which may seek to mould another state's public opinion, or disrupt elections, or increase animosity between political rivals and social groups, or altogether weaken democratic principles.

As opposed to traditional forms of propaganda, DFI is employed primarily through the Internet and specifically via social media platforms.

Russia's Internet Research Agency (IRA), for instance, ran digital campaigns across multiple social media platforms during the 2016 US presidential election. According to one report, this campaign resulted in 3841 persona accounts on Twitter generating 10.4 million tweets (of which 6 million were original), 81 unique Facebook pages containing 61,483 posts, 1107 videos across 17 YouTube channels, and 33 Instagram accounts containing 116,205 posts.

DFI campaigns also targeted Germany, the UK, France, and Taiwan. As outlined in this report, the process often starts with sophisticated hackers, often backed by a state, stealing sensitive personal and/or professional digital data. Next, the data are eventually dumped anonymously and made publicly available. Twitter and other social media platforms are then used to draw broader attention to the documents and data. Bots do their part to amplify the process even further. The content enters the collective mainstream, shared by regular social media users and reported upon by traditional media.

By manipulating and falsifying information, DFI threatens Canadian liberal democratic institutions in several ways: It replaces the interests of Canadians with those of a foreign government or actor; it inflames societal tensions within Canada, and within and among Canada's friends and allies, creating political polarization and division as a result; and it facilitates authoritarian overreach of Canadian citizens with connections abroad. More fundamentally, because of our increasing reliance on digital technology for information, DFI threatens to undermine the very notion of truth within Canadian society and degrade trust in the democratic process.

The threat posed by DFI is also evolving, making use of emerging technologies, including, most notably, Artificial Intelligence and Machine Learning (AI/ML). While contemporary DFI still has an important human element, in which information is generated and disseminated by people who publish material online via social media in much the same way as ordinary citizens might communicate their own political views to their friends and family, the future of DFI will be even further AI-enhanced and AI-generated.

Powerful automated software will troll the Internet, generating its own content and disseminating it against pre-selected and vulnerable populations. As artificial intelligence becomes increasingly sophisticated, the human element in DFI will diminish in importance. AI-supported software may eventually autonomously generate manipulative or suggestive photographs, videos, and text. While humans may retain command over the decision to launch a DFI campaign as part of a larger political or international strategy, the campaign itself might be executed and managed by an artificially intelligent software program.

Of particular concern are deepfakes, which are "video forgeries" that appear to make people say or do things they never did. With enough photo or video images of a person, facial recognition algorithms can recreate a solid replica of the person's original face. The material can then be superimposed onto other video content. Add audio – also facilitated by AI – and you have a convincing video of a person engaged in a scenario that never took place. The goal would be to produce and distribute fake but highly realistic and customized content to strategically shift narratives and perceptions, and, ultimately, behaviour.

Responding effectively to DFI will require a multifaceted, multilateral, and flexible approach. Internet companies and social media firms will have to be held accountable for the information they disseminate and post on their sites. These companies should be encouraged to find better ways to identify, flag, correct, and even eliminate false or misleading information in a timely and efficient manner. Independent tribunals might be established to review and possibly reinstate material that is removed. States may need to promote a common legal understanding of the phenomenon of disinformation and misinformation among and between the private and public sectors.

Canada should continue working with like-minded states to counter DFI when and where it occurs. Providing a common baseline for response and collective action will help individual democracies present a unified front. Working in partnership with others, Canada might cautiously explore whether and how it might use DFI against known and identified aggressors. For the future, Canada should encourage the continued private sector development of domestic AI excellence in a manner that finds the right balance with privacy rights. It can also explore ways to better integrate AI expertise into Canada's defence establishment.

Sommaire

L'interférence étrangère sous forme numérique (IEN) est un nouveau défi en matière de sécurité nationale au Canada qui progresse au fur et à mesure qu'un nombre croissant d'intervenants malveillants à l'étranger apprennent à exploiter efficacement cette tactique contre les démocraties. L'IEN est particulièrement dangereuse, car les stratagèmes employés par ces intervenants peuvent éroder de manière sournoise les institutions, les valeurs et les normes établies au moyen de contacts directs avec les citoyens étrangers par des voies de communication étendues et gratuites.

L'IEN est une forme évoluée de propagande traditionnelle qui utilise l'Internet et d'autres types de technologies pour créer et propager des renseignements de manière immédiate, ciblée et personnalisée. Contrairement aux formes traditionnelles de propagande, l'IEN mise principalement sur l'Internet, et plus particulièrement sur les plateformes de médias sociaux. Lorsqu'un individu malveillant noue un lien virtuel avec des personnes ou des collectivités étrangères, une propagande ciblée et personnalisée peut aisément être orchestrée et diffusée. Dans certains cas, cette propagande soutient les objectifs nationaux d'un gouvernement étranger, qui peut chercher à façonner l'opinion publique d'un autre État ou, encore, perturber des élections, aiguiser l'animosité entre rivaux politiques et groupes sociaux ou discréditer complètement les principes démocratiques.

Contrairement aux formes traditionnelles de propagande, l'IEN mise principalement sur l'Internet, et plus particulièrement sur les plateformes de médias sociaux.

Le Centre de recherche Internet russe (Internet Research Agency ou IRA), par exemple, a mené des campagnes numériques sur de nombreuses plateformes de médias sociaux lors de l'élection présidentielle de 2016 aux États-Unis. Selon un rapport, cette campagne a été à l'origine de 3 841 nouveaux comptes personnels sur Twitter – au moyen desquels 10,4 millions de gazouillis ont été transmis et 6 millions de nouvelles conversations ont été alimentées –, de 61 483 messages dans 81 pages individuelles sur Facebook, de 1 107 vidéos sur 17 chaînes YouTube et de 116 205 messages provenant de 33 nouveaux comptes sur Instagram.

Les campagnes de propagande ont également ciblé l'Allemagne, le Royaume-Uni, la France et Taïwan. Comme on le décrit dans la présente étude, le processus démarre souvent par un vol de données numériques personnelles ou professionnelles de nature délicate par des cyberpirates sophistiqués, souvent au service d'un gouvernement. Les données sont ensuite téléversées en vrac de manière anonyme pour être mises à la disposition du public. Puis, des conversations sont lancées sur Twitter et d'autres plateformes de microblogage pour attirer l'attention d'un large public sur ces informations. Les robots indexeurs font aussi leur part pour amplifier l'écho sur la Toile. Ces contenus, partagés dans les médias sociaux réguliers et repris par les médias traditionnels, font dès lors leur chemin dans l'esprit collectif.

En manipulant et en falsifiant les informations, l'IEN menace les institutions démocratiques libérales canadiennes de plusieurs manières : elle substitue les intérêts des Canadiens à ceux de gouvernements ou d'intervenants étrangers; elle aggrave les tensions sociales à l'intérieur du pays et parmi les amis et alliés du Canada, ce qui suscite la polarisation et la division sur le plan politique; et elle prédispose au dogmatisme les citoyens canadiens en contact avec l'étranger. Plus fondamentalement, en raison de notre dépendance croissante à l'égard de la technologie numérique comme source d'information, l'IEN risque d'affaiblir la notion même de vérité au sein de la société canadienne et d'éroder la confiance envers le processus démocratique.

La menace posée par l'IEN évoluera également, en misant sur les technologies émergentes, tout particulièrement sur l'intelligence artificielle (IA) et l'apprentissage machine. L'IEN contemporaine intègre toujours un élément humain important, car l'information apparaît et circule en ligne grâce aux messages transmis sur les médias sociaux, tout comme s'il s'agissait d'un face à face animé par des citoyens ordinaires qui confient à leur entourage leurs propres opinions politiques. Néanmoins, l'IEN de l'avenir tirera grand profit de l'intelligence artificielle en matière de création et de diffusion.

Un logiciel automatisé puissant explorera l'Internet en générant et en diffusant son propre contenu à l'intention de populations ciblées et vulnérables. À mesure que l'intelligence artificielle deviendra de plus en plus avancée, le facteur humain dans l'IEN diminuera. Les logiciels pris en charge par l'IA pourraient éventuellement générer de manière autonome des photographies, des vidéos et des textes faux ou suggestifs. Les humains seront toujours aux commandes pour lancer une campagne d'IEN dans le cadre d'une stratégie politique ou internationale étendue, mais la campagne elle-même pourrait être menée et gérée par un logiciel doté d'une intelligence artificielle.

Ce sont les « deepfakes », vidéos fabriquées de toutes pièces où des personnes parlent ou font des choses, qui suscitent le plus d'inquiétude. Avec suffisamment d'images photo ou vidéo d'une personne, les algorithmes de reconnaissance faciale réussissent assez bien à reproduire son visage. Un contenu vidéo peut ensuite être superposé à un autre. Ajoutons le son – étape également facilitée par l'IA – et nous obtenons une vidéo convaincante à propos d'une personne participant à un scénario qui ne s'est jamais produit. L'objectif serait de produire et de diffuser un contenu factice, mais très réaliste et personnalisé, afin de faire évoluer de manière stratégique les idées et les perceptions et, par conséquent, de favoriser certains comportements.

Répondre efficacement à l'IEN exigera une approche multiforme, multilatérale et souple. Les sociétés Internet et de médias sociaux devront être tenues responsables des informations qu'elles diffusent et affichent sur leurs sites. Ces sociétés doivent être incitées à trouver de meilleurs moyens de repérer, de signaler, de corriger, voire même de retirer les informations fausses ou trompeuses, et ce, rapidement et efficacement. Des juges indépendants pourraient être désignés afin d'examiner les documents retirés et de rétablir éventuellement leur statut. Les États pourraient devoir promouvoir une acception juridique commune pour désigner le phénomène de désinformation et de mésinformation entre les secteurs privé et public et à l'intérieur de ces derniers.

Le Canada doit continuer à collaborer avec des États partageant les mêmes valeurs que lui pour contrer l'IEN quand et où il se produit. Se rallier pour intervenir collectivement aidera les démocraties individuelles à présenter un front unifié. En partenariat avec d'autres pays, le Canada pourrait examiner minutieusement si et comment il pourrait exploiter l'IEN contre des agresseurs connus et identifiés. Pour mieux préparer l'avenir, le Canada devrait pousser le secteur privé à continuer de développer l'excellence en matière d'IA nationale tout en respectant le droit à la vie privée. Il pourrait également étudier des moyens de mieux intégrer les compétences en matière d'intelligence artificielle dans le système de défense canadien.

Introduction

Digital foreign interference (DFI) is an emerging Canadian national security challenge that is rising in significance as more malicious foreign actors learn how to effectively employ it against democracies. DFI is particularly dangerous because foreign actors can surreptitiously undercut established institutions, values, and norms by directly connecting with citizens in other countries through free and widely accessible methods of communication. As opposed to traditional forms of propaganda, DFI is employed primarily through the Internet and specifically via social media platforms. Once a malicious actor is virtually connected with foreign individuals and communities, they can create and disseminate tailored and targeted propaganda. In some cases, this propaganda feeds the national objectives of a foreign government, which may seek to mould another state's public opinion, or disrupt elections, or increase animosity between political rivals and social groups, or altogether weaken democratic principles.

Although individuals and states have attempted to intervene in foreign nations without resorting to military engagement for as long as statecraft has existed, the reach of online, digitized information has created novel dilemmas for liberal democratic institutions. By its very nature, transnational, digital information is vulnerable to interpretation, manipulation, and hacking by foreign actors outside of another state's jurisdiction. The cross-border nature of the Internet has therefore created unique challenges for democratic states in particular.

By manipulating and falsifying information, DFI threatens Canadian liberal democratic institutions.

By manipulating and falsifying information, DFI threatens Canadian liberal democratic institutions in several ways: It replaces the interests of Canadians with those of a foreign government or actor; it inflames societal tensions within Canada, and within and among Canada's friends and allies, creating political polarization and division as a result; and it facilitates authoritarian overreach of Canadian citizens with connections abroad. More fundamentally, because of our increasing reliance on digital technology for information, DFI threatens to undermine the very notion of truth within Canadian society and degrade trust in the democratic process.

The threat posed by DFI is also evolving, making use of emerging technologies, including, most notably, Artificial Intelligence and Machine Learning (AI/ML). While contemporary DFI still has an important human element, in which information is generated and disseminated by people who publish material online via social media in much the same way as ordinary citizens might communicate their own political views to their friends and family, the future of DFI will be even further AI-enhanced and AI-generated. Powerful automated software will troll the Internet, generating its own content and disseminating it against pre-selected and vulnerable populations. As artificial intelligence becomes increasingly sophisticated, the human element in DFI will diminish in importance; AI-supported software may eventually autonomously generate manipulative

or suggestive photographs, videos, and text (Allen and Chan 2017). While humans may retain command over the decision to launch a DFI campaign as part of a larger political or international strategy, the campaign itself might be executed and managed by an artificially intelligent software program. Humans need not apply.

This report makes sense of DFI's past, present, and future, with an eye towards protecting Canadian democracy. It begins by highlighting how and why DFI threatens Canadian democracy. It turns next to an exploration of recent examples of DFI campaigns that have targeted Germany, the UK, France, and Taiwan, illustrating how it functions in practice. The paper then turns to a discussion of how AI will change the nature, use, and utility of DFI in the coming years and decades. The paper concludes with thoughts on how Canada and its allies might better protect themselves from DFI.

DFI: Propaganda for the Internet Age

At its root, DFI is any deliberate action or attempt by foreign actors to clandestinely influence citizens of another state by, among other methods, presenting them with false information or misleading narratives by means of the Internet or any other digital platform. DFI has much in common with propaganda, which "seeks to dominate public thought processes so as to influence actions" (Baines and O'Shaughnessy 2013, xxi). Historically, propaganda was used by both state and non-state actors to change foreign attitudes and manipulate foreign political behaviour. Propaganda and DFI share the same underlying logic and objective. DFI is an evolved form of traditional propaganda, whereby the Internet and other types of technology are utilized to create and proliferate particular information in an immediate, targeted, and tailored way.

Today, over half of Canadians tune in online first to get their news (Zimmer and Erskine 2018). Methods of political advertising are evolving to reflect this trend. For example, during the 2016 American presidential election cycle, $1.4 billion was spent on online advertising, up from a mere $160 million in 2012 (Kim 2018, 1). This is a testament to the continuing shift in importance from physical news sources to digital ones. Despite these trends, however, digital media – in comparison to print, radio, or televised media – is not (yet) subject to the same level of legal and political scrutiny, regulation, or oversight. For instance, Canadian broadcasting and telecommunication regulations mandate that broadcasters and television stations not display abusive comment or images that are likely to expose an individual or a group to hatred on the basis of their identity, or any "false or misleading news" (Canada 2015, 2018d). Digital media face few similar restrictions. "We don't have a policy," Facebook argued in a May 2019 statement about fake videos appearing on its site, "that stipulates that the information you post on Facebook must be true" (Harwell 2019b).

States have only just begun grappling with adapting their laws and regulations for information created and disseminated online (Kim 2018, 3). Until they do, reality on the Internet will remain pliable, providing propagandists with avenues for sowing political discord (DiResta et al. 2017). Besides a shift in media consumption, Canadians are also increasingly migrating every facet of their lives online. The data have followed. From banking, education, and communications, to business and shopping, every online activity adds to the store of data. Digital information is by

its very nature transnational, making it vulnerable to manipulation and theft by foreign actors (Piccone 2018).

DFI threatens Canadian democracy in several ways. First, it can manipulate information so that foreign (rather than domestic) interests are reflected, expressed, and disseminated. Civic literacy is defined as the "knowledge and capacity of citizens to make sense of their political world" (Dixson 2005). Despite the fact that there are few empirical methods to test civic literacy levels, it is nonetheless considered a crucial element of liberal democratic society. If civic literacy is low, the ability of citizens to make rational and informed political decisions within the democratic process is compromised. DFI purposefully and negatively impacts democratic civic literacy because it is intentionally deceptive in nature.

DFI can create or promulgate political tension that can lead to civic discord and violence.

The content packaged in DFI ranges from outright false information – otherwise known as disinformation – to highly biased forms of narrative framing – also known as misinformation (Canada 2018c). False or highly biased information can reduce a democratic citizen's ability to properly contemplate and identify political decisions that serve the national interest. With DFI, the manipulation of facts compels a citizen to unintentionally reflect upon the interests of a foreign government or entity. For example, Russian DFI efforts have targeted Canadians by proliferating misinformation about Syrian refugees in order to sow discord and disagreement on the country's pro-refugee policies. Decreasing support for Syrian refugees in democratic countries, like Canada, may in turn play into Russian interests in the Middle East, by perpetuating its engagement in the region (Rocha 2018).

Second, DFI can create or promulgate political tension that can lead to civic discord and violence. Liberal democratic societies uphold human rights within and outside their borders (United Nations n.d.). Canada has long been a vocal champion of universal human rights. Disinformation and misinformation can undermine these efforts, exacerbating tensions between social groups, and in turn generating support for policies and actions that counter and weaken global human rights. As an illustration, a genocide was incited by the military in Myanmar through a concerted and multiyear effort that played out, in part, on social media: Targeted disinformation and misinformation was spread to legitimize violence against the country's minority Rohingya population (Mozur 2019). Brutal violence ensued.

Third, DFI can facilitate authoritarian government overreach within democracies, undermining the concept of universal freedoms within the democracy. For example, Canada is home to a sizeable number of citizens and permanent residents who have family in, or cultural, economic, and other ties to China. The ruling Chinese Communist Party (CCP) works to foster loyalty to the party through "pre-emptive social government processes," often facilitated by AI and other technology (Canada 2018b). China's new social credit scheme is currently being used by the CCP to expand the influence, control, and power of the government. This is problematic for

liberal democratic countries because the social credit system appears to operate well outside of China's physical borders (ibid.). In this case, the personal data of Chinese Canadians may be compromised even while they are physically removed from China and used to help tabulate their "social score" within China.

Finally, DFI threatens the notion of an established truth, of facts that are known and shared and acted upon. One possible effect of this is a reduction of public trust in democratic institutions. Research from MIT suggests that online misinformation and disinformation spread more quickly than factual information, that falsehoods have a greater reach than truths (Vosoughi, Roy, and Aral 2018). The reason may have something to do with the notion that contentious issues – often the very basis for misinformation and disinformation – elicit emotional human responses that compel people to engage with the material more passionately and forcefully (ibid.). Acquiring information, even useless or damaging information, feels good – literally; it has an effect on the "brain's dopamine-producing reward system," similar to money or junk food (Counts 2019). Researchers have also found that fake news travels faster and more broadly online than real news does (Vosoughi et al. 2018; Xiao et al. 2019).

All of this is problematic because the information provided in DFI often conflicts with the mainstream narratives espoused by traditional media and governments alike. Indeed, traditional media can either inadvertently provide legitimacy by reporting DFI information without proper corroboration or actually be co-opted by foreign actors in order to corroborate and legitimize electronic disinformation and misinformation. A good example is the capture of Chinese-language media in the Canadian market, which are owned by businesses with ties to the CCP (Blackwell 2019). The dissonance between fact and fiction can confuse citizens about which "facts" are correct and can help foster and bolster potentially harmful conspiracy theories. The spread of DFI is scientifically and empirically linked to growing public distrust in established liberal democratic institutions (Bennett and Livingston 2018, 122).

How DFI Works: The Human Touch

Digital foreign information is transmitted either by individuals using the same methods and platforms that ordinary people generally employ to publish and share information online, or automatically by bots – automated software designed to perform tasks on the Web. Currently, humans are still primarily principally responsible for DFI content generation, including the text and images transmitted online. Bots play a largely supportive role in amplifying the message across the Internet and various social media platforms, with the goal of reaching not only the largest but also the most relevant audience through microtargeting.

While DFI often appears esoteric and technologically advanced, the manner in which mis/disinformation is spread can be surprisingly simple. For example, Russia's Internet Research Agency (IRA) – commonly referred to as Russia's troll factory – ran digital campaigns across multiple social media platforms, including Facebook, Instagram, Twitter, and Youtube, which were designed to target US voters as a means of polarizing American political discourse and debate. This often entailed using a variety of "fake accounts" or creating "sock puppet" accounts "that mimicked a number of different types of legitimate users" (Howard et al. 2018, 25). Moreover, in their efforts to influence the outcome of the 2016 US presidential election, the IRA made use

of the same cross-platform digital advertising strategies that have been employed in the private sector to focus consumer attention on marketed products (ibid., 8).

The IRA's digital campaign is a good example of "computational propaganda," which relies on the use of "automation, algorithms, and big-data analytics to manipulate public life" (ibid., 39). Most IRA accounts on Twitter were automated, for example, although "they would frequently show signs of manual intervention" (Cleary 2019). In turn, regular Twitter users would often unwittingly retweet IRA posts, thereby amplifying the message and allowing IRA accounts to be assimilated into this foreign social media environment. Bots, automation, and big-data analytics do play an important and growing role in DFI. Yet it is important to note that current DFI campaigns make use of the same basic technologies that are available to average citizens on social media platforms (DFRLab 2019).

We also need to recognize the human element present within DFI. A good example is the so-called Chinese "50 cent party" – 50 cents being the alleged rate at which individuals are paid per post – which contains between 500,000 and 2 million Internet commentators. Reviewing leaked Chinese documents, Harvard University's Gary King and colleagues found that "each [online post] was written by a specific, often identifiable, human being under direction from the government" (King, Pan, and Roberts 2017, 489). Given that the 50-cent party is estimated to be responsible for approximately 450 million annual postings online, the sheer scope of this endeavour is evident (ibid., 485). Perhaps surprisingly, only about half of the online posts were made on commercial sites, such as Sina Weibo, Tencent Weibo, and Baidu Tieba, popular Chinese microblogging and communications platforms (ibid., 488). The remaining posts were made on Chinese government sites. Equally important is the role of instant messaging apps, such as WeChat or Line, to spread dis/misinformation. Unlike Russian Internet trolls, who often focus their energy abroad, Chinese commentators appear more focused on "cheerleading" government policies, with only a tiny minority engaged in argumentative or taunting posts (ibid., 490).

Chinese commentators appear more focused on "cheerleading" government policies.

Similarly, although bots played a significant role behind Russian DFI, attention should also be paid to the role of people who took part in this campaign. Leaked documents released by hackers in 2013 revealed that "Russia's Internet Research Agency employed over 600 people with an annual budget of $10m" (Benedictus 2016). Regardless of whether the particular employee was assigned to comment on news articles or maintain falsified Facebook or Twitter accounts, each human troll was responsible for producing a given amount of posted material each day. A quota had to be met: "All were assigned specific targets and goals for the number of followers they needed to attract" (Bennett and Livingston 2018, 132).

The amount of IRA activity across the various social media platforms was staggering. The process was documented by American researchers working with New Knowledge, an "information integration company" that works to protect brands, actors, and industries from online campaigns that manipulate narrative and "damage reputation." On behalf of the US Senate Select Committee on Intelligence, which provided New Knowledge with an "expansive data set" of

Table 1: IRA Activity across Social Media Platforms[1]

PLATFORM	CONTENT
YouTube	**1107** videos across **17** YouTube channels
Twitter	**3841** persona accounts on Twitter generating **10.4 million** tweets (of which 6 million were original)
Facebook	**81** unique Facebook pages containing **61,483** posts
Instagram	**133** Instagram accounts containing **116,205** posts

Source: DiResta et al. 2018.

social media posts and metadata collected from Facebook, Twitter, Alphabet, and other online platforms, the organization was able to detail how the IRA's "influence operations targeting American citizens" between 2014 and 2017 took place. A tabulated breakdown of New Knowledge's findings is provided in table 1.

While the content shared on these platforms primarily contained memes, images, and short messages and posts, the IRA was also engaged in producing more elaborate disinformation. For example, Vitaly Bespalov, a former employee in one of Russia's misinformation entities, stated that his assignment was to rewrite news articles about the war in eastern Ukraine, "to chang[e] the text in order to give articles the appearance of originality and a distinctly pro-Russian slant" (Maynes 2018). Bespalov's goal was not to invent or fabricate stories, but rather to depict and reshape actual events to support the Kremlin's objectives in Ukraine: "We'd switch the word 'annexation' of Crimea for 'reunification,' or call the government in Kiev 'a fascist junta' while writing favorably about the separatist rebels in Eastern Ukraine" (ibid.). Thus, alongside the relatively unsophisticated content espoused and disseminated by perpetrators of DFI, which is designed to generate an emotional response from the audience, foreign governments have likewise engaged trained specialists to clandestinely reframe content in order to manipulate public opinion. These tactics are explored further and in greater detail in the following section, which highlights DFI campaigns targeting Germany, the UK, France, and Taiwan.

Targeting Democracy: DFI in Action

To better appreciate how DFI has been employed, and to what end, an exploration of recent DFI campaigns launched against four of Canada's allies and partners follows.

Germany: Tearing the Heart out of Europe

Although the 2017 German Federal elections experienced a general lack of direct political interference compared to what other European states faced in their own recent elections, Germany has not been immune to DFI (Stelzenmuller 2017). In many ways, Germany, which has become the political and economic centre of Europe and is the primary force behind Europe's collective response to the migrant crisis, is a primary target of DFI. Russia, in particular, appears to have Germany in its sights.

Russia's goal is less about stealing or hacking a German election or purposefully propping up a favoured, and presumably pro-Russian, candidate, but rather to muck up the German electoral system itself.[2] "Cyberattacks are taking place with no other motive than to cause political uncertainty… to put pressure on open discourse and on democracy," explains Bruno Kahl, head of Germany's Federal Intelligence Service (*Bundesnachrichtendienst*). "Europe," Kahl adds, "is now the focus of these interference attempts, and Germany in particular. The perpetrators are mostly interested in delegitimizing the democratic process, no matter whom they help get ahead" (King 2016). Several German examples illustrate how the process works.

In 2015, for instance, hackers with ties to the Russian government targeted the Bundestag.

In 2015, for instance, hackers with ties to the Russian government targeted the Bundestag – specifically, Chancellor Angela Merkel's Christian Democratic Union (CDU), as well as the German Ministry of Finance and the Ministry of Foreign Affairs (Brattberg and Maurer 2018). This attack entailed hacking into individual email accounts, leaking the information through proxies such as WikiLeaks, and then utilizing bots, human trolls, and state-controlled media outlets to spread and amplify the disinformation. German intelligence sources reported that the source of this attack – an actor dubbed APT28 (Advanced Persistent Threat 28) or Fancy Bear – was the same Russian cyber espionage organization that reportedly hacked the US Democratic National Committee during the 2016 presidential election, alongside the 2017 French elections (ibid.). In all three cases, hacked information was rolled into a larger DFI campaign, which involved leaking both factual and at times fabricated information to the public.

Another prominent German case of DFI surrounds the alleged disappearance and rape of a 13-year-old Russian-German girl in January 2016, commonly known as "The Lisa" case. The girl – who police later determined had been visiting and staying at a friend's house – was thought to have gone missing over a 30-hour period (Meister 2016). Picking up on the girl's false claims

that she had been sexually assaulted during that period by foreigners and migrants, the story was extensively covered by domestic and foreign media, culminating in a diplomatic spat between Germany and Russia.

Over a period of several weeks, several Russian-language outlets, including foreign media such as RT, *Sputnik*, and RT Deutsch, reported on the case; anti-migrant right-wing groups posted material on social media; and public demonstrations were organized via Facebook by German-Russian and far-right organizations (Meister 2016). At one point, Russian Foreign Minister Sergei Lavrov accused Germany of a cover-up: "The news of her disappearance was kept secret … They [German authorities] are painting over reality with political correctness" (Huggler 2016). The end result was a firestorm of intensive anti-migrant sentiment within Germany, whipped up by highly emotional political rallies and amplified through social media.

United Kingdom: Sowing Division, Mistrust, and Confusion

Like Germany, recent DFI campaigns aimed at the UK have primarily originated from Russia. A London School of Economics paper revealed that the Russian government's aims in the UK are based on promoting its foreign policy agenda through local proxies (including among far-right and pro-Brexit organizations), eroding trust in democratic institutions, increasing political polarization, and spreading confusion during crises, elections, and other key national events (Applebaum et al. 2018, 3). Another report, published by the Oxford Internet Institute, noted that the UK has been a target of DFI around national elections, with the issue of immigration and migration often taking centre stage. And, the campaigns have been enabled by both bots and human social media accounts (Bradshaw and Howard 2018, 14).

The result of these and other DFI campaigns in the UK are difficult to measure, however. While the British House of Commons, in an October 2018 Interim Report, *Disinformation and 'Fake News*,' concludes that "the Government has not seen evidence of successful use of disinformation by foreign actors, including Russia, to influence UK democratic processes," subsequent reports by the House of Commons suggested otherwise (ibid., 16). In its February 2019 Final Report, it concluded that defining successful interference "is impossible" and that there was otherwise "strong evidence" from third-party experts and research institutions that "point to hostile state actors influencing democratic processes" (UK 2019).

Examples of DFI from the UK include several notable events. For instance, DFI targeted the results of the 2014 Scottish Independence Referendum, in which pro-Kremlin social media accounts later distributed fake video content meant to discredit and subvert the integrity of the vote count itself (Carrell 2017; DFRLab 2017b). The content was curated to depict vote rigging, with the intention of inflaming supporters of Scottish independence, who narrowly lost.

Later, in 2017, on the heels of a string of deadly terrorist attacks in London and Manchester perpetrated by supporters of the Islamic State, Russian-based social media content was again used to spin the attacks in a way that could inflame anti-migrant and anti-Muslim sentiment in the UK and Europe more broadly. Researchers at the Centre for Research and Evidence on Security Threats – a consortium that brings together major British universities – found that 47 different social media accounts were used to "influence and interfere with public debate following" the attacks. Eight of the accounts posted nearly 500 Twitter messages, which were reposted over 150,000 times. Some of the accounts were broadcasting "inflammatory messages" within 15 minutes of the attacks, attempting to "frame" the events in a certain light. One such tweet, published within one hour of a suicide bombing in Manchester – "Another day, another Muslim terrorist

attack. RETWEET if you think Islam needs to be banned RIGHT NOW!" – was reposted nearly 4,000 times (CREST 2017, 2).

And, in 2018, following the botched assassination, by poisoning, of an ex-Russian spy, Sergei Skripal, on British soil by two Russian agents, there was a concerted online campaign initiated by Russian sources that sought to undermine and ridicule the British government's investigation of the attack (Applebaum et al. 2018). When UK authorities released CCTV images of the two Russian suspects passing through two gates upon arriving at Gatwick airport outside London, pro-Kremlin social media accounts immediately pounced on the fact that the images had identical timestamps: They ridiculed UK authorities as a result, painting the event as a British intelligence operation.

Russian Foreign Ministry spokesperson, Maria Zakharova, then chimed in: Either "the date and the exact time were superimposed on the image," or the supposed Russian spies had "mastered the skill of walking simultaneously" (Gunter and Robinson 2018). It did not ultimately matter that Gatwick has multiple, identical non-return gates, and that the two suspects could easily have passed through them at the same time if they were disembarking together. What mattered was that people were reframing and questioning the investigation, putting into doubt the veracity of the British government's version of events.

According to Ben Nimmo, an expert on Russian disinformation: "What is really striking is that you no longer see the Russian machine pushing a single message, it pushes dozens of messages. The more different theories you put out, the more different Google results you're going to get … and for someone who is not following the story regularly, that becomes more and more confusing" (Gunter and Robinson 2018). New conspiracy theories surrounding the British investigation soon emerged online, including that the suspects were hired British actors or MI6 agents, and humour was used to poke fun at British authorities and mock their investigation.

When the UK named the Russian suspects, pro-Kremlin accounts posted near-identical tweets, for instance, with pictures of famous Hollywood spies and historical figures – like Jason Bourne and Joseph Stalin – instead of the suspects. And once again, Russian officials then chimed in: Russia's embassy in London posted content on its social media page that juxtaposed the image of the suspects alongside British authorities in biohazard suits, asking readers to "spot the difference" (ibid.). The British House of Commons concluded in 2018 that Russia was seeking to "weaponise information" to achieve its strategic goals, and that in the specific case of the Skripal poisoning, it had "promulgated at least 38 false disinformation narratives" to do so (UK 2018).

France: New-age Vote Rigging

On May 5, 2017, two days before the second round of the French presidential election, a large number of files were uploaded to the anonymous sharing site PasteBin, titled "EMLEAKS" (Mohan 2017). The files contained thousands of emails – some real, some fabricated – and other political documents related to *En Marche!*, a political party founded by Emmanuel Macron, then the election's frontrunner. The document release was timed to coincide with the 44-hour media blackout that precedes all French elections – an election rule that dates back decades – presumably in order to prevent Macron from countering any pejorative content contained in the documents.

Nearly two years after the hack, the French government has yet to officially blame any country or individual for the political interference, noting only that, "What can be safely assumed is that, whoever the perpetrator was, they were at least linked to Russian interests and received help from the American alt-right and French far right, two communities that share a very close vision

to that which is articulated by the Kremlin" (Vilmer et al. 2018, 111). Though Macron ultimately won the election, the so-called Macron Leaks highlight the process by which the DFI campaign was orchestrated.

Although the document dump garnered the most public attention, the document release was only the culmination of a much longer and internationally coordinated disinformation campaign orchestrated against Macron (Vilmer 2019). First, several political institutions and parties were hacked in order to illegally secure sensitive and potentially damaging information (Toucas 2017). At least five of Macron's closest aides and allies had their professional and personal emails hacked, including his speechwriter, the campaign's treasurer, and two MPs (Vilmer et al. 2018, 107).

The so-called Macron Leaks highlight the process by which the DFI campaign was orchestrated.

The hacks were followed by a "diffusion of rumors and insinuations" published by various media between January and February 2017: On February 4, 2017, for example, an article by *Sputnik* (2017) – a tabloid linked to the Russian government (Nimmo 2016) – presented Macron as a "US agent supported by a very wealthy gay lobby." Then, just hours before the final leaders' debate between Macron and Marine Le Pen – leader of France's far-right political party, National Rally (formerly, National Front) – on May 3, 2017, two fake documents alluding to Macron's secret offshore accounts were posted to 4Chan – an online, English-language bulletin-board (of sorts) that allows users to post and share material anonymously. The post was then disseminated via Twitter by 7000 accounts, with the #MacronGate hashtag. During the debate itself, Le Pen mentioned the "existence of hidden accounts" herself (Vilmer et al. 2018, 107). Finally, days later and just hours before the stolen documents were ultimately leaked, rumours of sensitive political documents began circulating the Web, not least by Julian Assange of WikiLeaks, who claimed to have "interesting documents" about Macron (Toucas 2017). The suspense surrounding the pre-election document leak was carefully built-up, with the intent of causing maximum damage to Macron.

A key takeaway from the Macron Leaks is that while the initial cyberattack was most likely orchestrated by a foreign power – i.e., Russia – the actual DFI campaign itself and the spreading of misinformation were carried out by a consortium of French and foreign groups and individuals. Indeed, the hashtag #MacronLeaks was coined by Jack Posobiec, an American alt-right commentator based in Washington, DC, who was one of the first people to share links to the stolen documents. #MacronLeaks reached over 45,000 tweets in under four hours (Volz 2017). Presumably, Mr. Posobiec has no personal stake in the outcome of the French election, other than an ideological affinity with Le Pen and her political movement. Only after the American far right popularized the document dump, did news about the leaks migrate to France, where it was then picked up by Le Pen supporters (Chhor 2017). Bots were also used to further disseminate material: The ten most active accounts using #MacronLeaks posted over 1,300 tweets in just a few hours, a "classic sign of bot use" (DFRLab 2017a).

Taiwan: Targeting Public Opinion and Independence

Chinese leaders view the Internet as both a threat to the stability and continued rule of the Communist Party, as well as a tool for population control and dissemination of propaganda (Canada 2018a). Although the goals of China's digital strategy extend to many areas, leaked documents published by the *China Digital Times* – an online news platform banned in mainland China – reveal that online Chinese commentators were directed to influence Taiwanese public opinion and democracy. Some of the methods outlined include making the United States the target of criticism, playing down the existence of Taiwan, explaining how democracy is not well-suited to capitalism, showing how the West is forcibly pushing its values abroad, and stirring up "pro-Party and patriotic emotions" (Xiao 2011). More than ever, DFI now permits China to target individual Taiwanese citizens and more covertly discredit the notion of an independent Taiwan in world opinion.

In recent years, China's interference with Taiwanese politics has changed from military intimidation during Taiwanese elections to information and economic warfare (Tsai 2018). In 2014, Ko Wen-je successfully ran for mayor in Taipei, backed by pro-independence Democratic Progressive Party (DPP), against Sean Lien, a more seasoned Nationalist Party (KMT) politician. Ko Wen-je's campaign was considered transformative for Taiwanese politics in its use of social media and other digital platforms (Monaco 2017). For example, his campaign created and used a program able to "crawl public pages on Facebook and collect all the data it could—how many likes or shares each post got and how many people liked or followed the page, etc." (ibid., 11). However, as Taiwan's politicians became aware of the power of digital campaigning, there is considerable evidence that China has increased its own DFI against Taiwanese democracy.

DFI now permits China to target individual Taiwanese citizens.

After the election of pro-independence candidate Tsai Ing-wen as president in 2016, China launched the Diba Facebook campaign to "show the reaction of Chinese citizens to Taiwan's election of Tsai Ing" (ibid., 23). The campaign involved posting pro-mainland messages on Ing-wen's Facebook page, which eventually garnered nearly 50,000 comments and replies. The messages expressed opposition to Taiwan's independence, and extolled mainland China and the virtues of socialism (ibid.). While the intent behind the Diba campaign is unknown, one can speculate that flooding Ing-wen's Facebook page after she had just won an election on a pro-independence platform was meant to subtly remind her, and her followers, that China was still adamantly opposed to Taiwanese independence. Commentators have also expressed concern about China's intervention in Taiwan's 2018 local elections of city mayors and county and village leaders (Horton 2018).

In addition to interfering in Taiwan's elections, a Taiwanese government task force, established to monitor the spread of fake news on social news, has "unequivocal evidence" that "the Chinese government is using online content farms to create fake news to manipulate Taiwanese public opinion and polarize society" (Chien, Chung, and Chin 2018). Content farms (also known as content mills) refer to Internet platforms that employ large numbers of freelance

writers to create unlimited content (Tung 2016). Many are located in mainland China and are used to spread misinformation, including fabricated stories, about Chinese military exercises, the cross-Strait relations, and Taiwan's failed economic and agricultural policies. Much of this content is generated by people, though the material itself is spread and replicated online by bots. Of note, "every day, as many as 2,400 distinct pieces of disinformation targeting Taiwan originate on Facebook" (Cole 2018).

Perhaps the most notorious example of Chinese disinformation targeting Taiwan is the 2018 fabricated story that China's consular services in Osaka, Japan, dispatched buses to rescue Taiwanese citizens trapped at Kansai International Airport, Japan, following Typhoon Jebi (Ko 2018). The implication of this story was that Taiwan was either incapable of helping its own citizens or unwilling to do so, whereas China would always assist fellow nationals. In fact, China never dispatched any aid to the stranded Taiwanese – the airport authorities provided the transportation. But that did not stop mainstream media outlets in Taiwan, and elsewhere, from picking up the story without verification. Criticism then emerged, blaming Taiwan's office in Osaka for failing to assist the stranded tourists. As a result of this blistering criticism, Taiwan's official representative in Osaka, Su Chii-cherng, committed suicide.

China's campaign of subversion, at least with respect to the digital aspect, is very similar to Russia's more notorious actions against the United States, France, Germany, and the United Kingdom. Moreover, China's attempt to discredit Taiwan's success and independence also bears close resemblance to Russia's actions to portray the Baltic states and Ukraine as a "failed experiment of both post-Soviet transformation and Euro-Atlantic integration" (Canada 2018a, 36). It is important that Chinese news about Taiwan is viewed with the same skepticism that citizens in Western democracies regard news emanating from dubious Russian sources such as *Sputnik* or *RT*. Although the West has been focused on countering Russian digital interference, this illustrative case study highlights the importance of developing flexible policies that can respond to interference from wherever it emanates.

In sum, these episodes of DFI from Germany, the UK, France, and Taiwan highlight a particular pattern of operation, repeated in whole or in part across the different examples. The process often starts with sophisticated hackers, often backed by a state, stealing sensitive personal and/or professional digital data that might embarrass or otherwise threaten foreign political adversaries if publicly released. At times, especially inflammatory but fabricated data are added to the stolen data to augment their overall effect if and when they are released, or to sow general confusion. Next, the data are eventually dumped anonymously and made publicly available on a popular and free discussion board or via another intermediary, such as WikiLeaks. Twitter and other social media platforms are then used to draw broader attention to the documents and data, which are picked up and further disseminated by particular social media personas and online political communities. Bots do their part to amplify the process even further. This creates even broader popular awareness, and eventually, the content enters the collective mainstream, shared by regular social media users and reported upon by traditional media.

As far as DFI has evolved over the past several years, the future of DFI poses yet another set of challenges. The automation of the process, augmented by emerging disruptive technology such as AI, will help broaden the malicious use and possible effect of DFI on democracies. The following section explores the nexus between AI and digital foreign interference.

DFI 2.0: The Robotic Touch

Although certain bots perform benign, even useful, functions, such as indexing web pages – a capability utilized by all online search engines – bots have become a key component in DFI. Bots can perform very simple functions to spread information online, such as providing prompts for re-watching a video on YouTube to boost its overall view count, retweeting messages, and liking Facebook pages. Social bots can also generate very simple messages, which can be self-generated or pre-fabricated by humans. However, with further advances in artificial intelligence, bots can become increasingly sophisticated, capable of autonomously generating their own realistic messages and interacting with (unwitting) real users. As AI becomes progressively more sophisticated, it will be able to generate photographs, video, and texts that will increasingly become indistinguishable from similar material created by humans.

Consider synthesized text. Using open-source "text generation software" and training it against a dataset of UN General Assembly speeches offered since 1970 (also available free, online) researchers were able to train a machine to pump out fake but highly realistic UN political speeches. The resulting speeches ranged from the familiar and diplomatically mundane – "Climate change continues to be a major concern of all world leaders, and the Government of Mali reiterates its call for the UN to urgently adopt a series of measures …" – to the hateful and inflammatory – "Refugees are terrorists and are taking the lives of their citizens. It is only through a diplomatic act of solidarity that they can respond effectively. It is a humanitarian duty. Every effort must be made to ensure that the safety of all the world's civilians is not threatened" (Bullock and Luengo-Oroz 2019). The fabricated speeches aren't perfect, but at a cost of less than $10 and 13 hours of labour, they are also a bargain. Malicious actors could easily circulate these speeches online, tucked in among other legitimate UN material, as part of a targeted DFI campaign to smear a given leader or country.

Another similar example of the growing potential of AI is OpenAI's language model, GPT-2, which can "generate realistic text in a variety of styles, from news articles to fan fiction, based off some seed text" (Lowe 2019). Feed the machine some starter information, and it fills in the gaps to establish a coherent, written sample. For example, GPT-2 was able to create a coherent news story about the discovery of unicorns based on three lines of seed text (Mak 2019):

> Seed Text:
> *In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*
>
> GPT-2 Produced Response:
> *The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science. Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved. Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow. Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.*

Again, the generated text is not quite perfect. The language and phrasing is at times slightly off. Through repeated analysis, however, AI will continually strive to optimize the content of its messages for maximum rhetorical effect: AI's ability to persuade individuals will become increasingly potent the more data are available to it. OpenAI concluded that its algorithm was sufficiently dangerous that it chose to release only a "much smaller version of the model and withhold the data sets and training codes that were used to develop it" (ibid.).

Google's Duplex, a rival to Amazon's Alexa and Apple's Siri, provides a third impressive and recent example of fabricating material, but this time in the form of real-time oratory. Duplex is said to have passed the Turing Test[3] in spring 2018 after the machine was able to book appointments and make reservations with humans over the phone, live (Nieva 2018). Human receptionists on the other end of the line did not know or realize that they were speaking to a machine. Duplex "uses verbal ticks like 'uh' and 'um'," to pass itself off as a human: "It speaks with the cadence of a natural human voice, pausing before responding and elongating certain words as though it's buying time to think" (ibid.).

Of greatest concern, however, is the fact that just as DFI encompasses more than text or written material, and includes the use of images and videos, the future of AI-generated misinformation and disinformation will also extend to these media formats. Welcome to the world of deepfakes. A portmanteau of "deep learning" (one method of machine learning, which is itself a subset of AI) and "fake", deepfakes are "video forgeries" that appear to make people say or do things they never did (Solsman 2019).

DFI encompasses more than text or written material, and includes the use of images and videos.

The premise is rather simple. With enough photo or video images of a person, facial recognition algorithms can unpack every minute detail and, over time and making millions of iterations, recreate a solid replica of the person's original face (Leslie, Hoad, and Spraggon 2018). The material can then be superimposed onto other video content. Add audio – also facilitated by AI – and you have a convincing video of a person engaged in a scenario that never took place. The process can be used on just about anyone: Most people are prolific posters, happily plastering images of themselves on social media. Target the right individual, however, and the technology is ripe for use in DFI. Writing for the Council of Foreign Relations, Robert Chesney and Danielle Citron (2018) argue that "the array of potential harms that deep fakes could entail is stunning": Well-crafted and well-timed deepfakes could "tip an election, spark violence … bolster insurgent narratives about an enemy's supposed atrocities, or exacerbate political divisions in a society."

Alina Polyakova, writing for the Brookings Institution, focuses on AI as a tool of asymmetric war; she calls it "AI-driven asymmetric warfare" (ADAW). With Russia in mind, Polyakova (2018) illustrates how weaker adversaries might "co-opt existing commercially available" AI technology to challenge stronger states with AI-generated disinformation or political influence campaigns that rely on deepfakes. The goal would be to produce and distribute fake but highly realistic and customized content to strategically shift narratives and perceptions, and, ultimately, behaviour.

Hyper-realistic synthetic videos might show President Donald Trump suggesting to Chinese President Xi Jinping, on the sidelines of an international gathering, that the US is willing to quietly let China increase its influence over Taiwan. Or, they might depict military leaders – say from Israel, France, Canada, or Saudi Arabia – uttering toxic statements about an ongoing military engagement – in Gaza, Mali, Iraq, or Yemen – that are tailored just right to inflame local resentment. Or, deepfakes might be used on the eve an election, depicting a top candidate as saying racist, misogynist, or otherwise harmful things (Dem 2019). Or, they could be used to coerce or threaten a politician: Think of Russia surreptitiously threatening a Ukrainian presidential nominee with engineered content that could influence the candidate's standing among the electorate. The possibilities are nearly endless (Tillman 2019).

Deepfakes are still imperfect: Glitches are often apparent to the naked eye. But the technology is improving every day (Piper 2019). In May 2019, for example, researchers published work on recent technological developments that bring still photographs and even portraits – of the Mona Lisa or Albert Einstein, for example – to life, using the portrait itself and data from real faces that closely resemble it (Zakharov et al. 2019). The result is stunning "living portraits" (Zakharov 2019). Step outside the computer science world and it is clear that creating nearly flawless doctored video and photo content can already be accomplished quickly and relatively cheaply. Do-it-yourself tutorials provide step-by-step instructions (Zucconi 2018).

Deepfakes are still imperfect: Glitches are often apparent to the naked eye.

For those with more money than time, deepfakes for hire are a popular option (Jianguo123 2018). At the aptly titled *thispersondoesnotexist.com*, a new, realistic, AI-generated face of a person that never lived is generated every time you refresh the page. These synthetic pictures might prove useful for establishing sophisticated fake social media accounts, as appears to have been the case featuring "Maisy Kinsley," a supposed Bloomberg journalist, equipped with a serious looking headshot, a Twitter bio, LinkedIn page, and personal website, fishing for information on social media (Fleishman 2019). And deepfake apps are available, free, online. One comes with a friendly disclaimer: "FakeApp was created to give everyday people access to realistic faceswapping technology for creative and exploratory use. If you use this tool, please use it responsibly" (deepfakeapp 2018).

Probably the most famous FakeApp forgery has former President Barack Obama suggesting that "President Trump is a total and complete dipsh**." Buzzfeed, an American media and entertainment company, created the video using American actor and filmmaker Jordan Peele – standing in for Obama – and 56 hours of an employee's time (Silverman 2018; Mack 2018). The video is meant as a Public Service Announcement, to warn the public of new-age fake news; it has been watched millions of times over.

Because determining the veracity of AI-manipulated content and attributing its source is difficult to do, countering this type of coercive misinformation may be challenging (Knight 2018). And, as deepfake technology becomes more commonplace, it will make us question the veracity of all video

content, both real and fake. That will allow individuals who were legitimately caught on tape doing or saying outrageous things to simply deny it (Leslie et al. 2018). Truth, once again, will suffer.

In a harbinger of how fake video content might spread online and influence political behaviour, consider the videos of prominent Democrat and speaker of the US House of Representatives, Nancy Pelosi, that appeared in May 2019. Originally taken at an event at the Center for American Progress, one video is subtly altered such that it appeared that Pelosi was slurring her words (*Guardian* 2019). The edited video was then posted on social media, where it was shared, re-posted, and otherwise viewed millions of times. Other, similarly altered clips of Pelosi eventually found their way to mainstream media – *Fox News* spent a segment dissecting the videos and commenting on Pelosi's mental and physical health (*NowThis Politics* 2019) – which were subsequently shared by President Donald Trump, no less. "PELOSI STAMMERS THROUGH NEWS CONFERENCE," President Trump tweeted in all caps (Musil 2019; Harwell 2019a). An analysis of the original and altered videos found that the latter had simply been slowed down, their pitch altered, to make it appear as though Pelosi was ill, or perhaps inebriated. The videos had been doctored. Their political effect – embarrassing Pelosi, galvanizing Trump's political base, and otherwise clogging the news cycle – was real.

The implication of all of this emerging technology and the potential dangers are clear: Instead of relying on thousands of humans to post original material, perpetrators of DFI could, eventually, rely on thousands of sophisticated algorithms and bots to create and spread nearly perfect fake or misleading information and news, engage in "live" social media disputes with human opponents, and otherwise generate mountains of disinformation at incredible speeds. And because the content will appear human-generated, organizations and individuals will have an increasingly difficult time distinguishing it from actual, online discourse.

Some organizations are turning to AI to counter AI-enabled DFI.

As a result, some organizations are turning to AI to counter AI-enabled DFI. With the release of GPT-2, researchers at MIT began developing counter software, GLTR, for instance, which can provide users with a probability score of whether a given text was machine- or human-generated (Mahmood 2019). Something similar is taking place with deepfakes: Companies and research institutions are developing tools to detect the coming generation of sophisticated, AI-generated forgeries (*Liwaiwai* 2019). Even the US military, via DARPA's Media Forensics program,[4] is supporting efforts to develop counter technologies that can automate the integrity assessment process of online images and videos (Villasenor 2019).

Similar to the situation emerging with text-based disinformation, an ongoing race is emerging between technologies that are capable of generating fake content and technologies that are capable of detecting it. AI will be central to both processes. The eventual end result is that countries with more advanced AI capabilities may be able to both disseminate and counter false information more easily than countries with less sophisticated AI. A country's ability to resist DFI in the future, and protect its democratic institutions, may be a function of its ability and willingness to develop and deploy AI defences.

Conclusion: Thoughts on the Way Ahead

Foreign disinformation campaigns have come a long way since the era of political propaganda. Technology has driven the process. Digital data have helped establish a repository of sensitive information that has proven difficult to protect and defend. The Internet, and the social media applications built upon it, provides individuals, communities, organizations, and countries with new ways to interact and share information in ways that circumvent traditional borders. And developments in software and AI provide the context for machine-generated content equal in quality to and greater in quantity than human-generated content.

Some states have proven willing and able to take advantage of these technological advances to develop tactics and strategies to surreptitiously turn real and fabricated information into political influence and power. The concern to Canada and like-minded democracies, whose societies and political systems rest on the free exchange of ideas and open political debate, is that these technological and strategic innovations, when wielded in certain ways, threaten to undermine the very social and political fabric upon which democracy rests.

Responding effectively to DFI will require a multifaceted, multilateral, and flexible response.

Responding effectively to DFI will require a multifaceted, multilateral, and flexible response. Internet companies and social media firms will have to be held accountable for the information they disseminate and post on their sites. These companies should be encouraged to find better ways to identify, flag, correct, and even eliminate false or misleading information in a timely and efficient manner. Independent tribunals might be established to review and possibly reinstate material that is removed, ensuring freedoms of online expression are upheld. Likewise, states should explore ways to persuade firms to be more transparent about how their algorithms function, encouraging them to adjust algorithms to minimize the promotion and reach of disinformation. That may require states to more clearly define fake news, disinformation, misinformation, and related terms, so as to promote a common legal understanding of the phenomenon among and between the private and public sectors.

From a geopolitical perspective, Canada should continue working with like-minded states, through the G7's Rapid Response Mechanism (RRM) and other similar endeavours, to counter DFI when and where it occurs (Canada 2019). Providing a common baseline for response and collective action will help individual democracies present a unified front. Research on threats, counter measures, and best practices should be collected and shared among allies, and joint training exercises should be held to counter DFI to better hone skills and responses. Non-G7 democracies should be invited to participate in these training simulations. Canada, working in partnership with others, might likewise explore – cautiously – whether and how it might use DFI against known and identified aggressors, if only to communicate a willingness to retaliate in kind to deter further nefarious engagement.

Finally, in terms of AI and the future of DFI, Canada should encourage the continued private sector development of domestic AI excellence in a manner that finds the right balance between providing researchers with access to appropriate and necessary training data – including, potentially, Canadian government data – and privacy rights. In a similar vein, Canada should explore the feasibility of developing shared norms and agreements with like-minded allied states that would permit the pooling of national data for use in collective AI research. If there is one undeniable advantage some states – notably, China – have over others, including the US, Canada, and the Europeans, with respect to AI development, it is data. Data fuel AI: The more (good) data at hand, the better the AI. And China is awash in them. According to Kai Fu Lee, author of *AI Superpowers: China, Silicon Valley, and the New World Order* (2018), China's "data edge is three times the US based on mobile user ratio, 10 times the US in food delivery, 50 times in mobile payment, and 300 times in shared bicycle rides."

Matching China's advantage will require pooling data resources multilaterally. This is already happening in Europe, with Germany and France seeking methods of bilaterally sharing their data in order to compete commercially with American and Chinese companies (Dittrich 2018, 2). Broader European cooperation will follow (EU 2018). If democracies are to match China's data advantage, Europe, the US, Canada, and others must work together, if not for commercial purposes, then at least for national defence. With a larger data pool and more skilled engineers and developers, Western democracies would be in an excellent position to defend against future, AI-enabled DFI campaigns. Finally, because DFI relates to statecraft and intelligence, Canada should explore ways to better integrate AI expertise into Canada's defence establishment: Private sector insights should be better leveraged within Canada's national security apparatus so as to assist the government in developing and deploying AI to counter future iterations of DFI.

About the Authors

Alex S. Wilner (https://alexwilner.com/) is an Assistant Professor of International Affairs at the Norman Paterson School of International Affairs (NPSIA), Carleton University and a Munk Senior Fellow at the Macdonald-Laurier Institute. He teaches classes on terrorism, intelligence, international affairs, and Strategic Foresight. Professor Wilner's research primarily focuses on the application of deterrence theory to contemporary security issues, like terrorism, radicalization, organized crime, cybersecurity, and Artificial Intelligence.

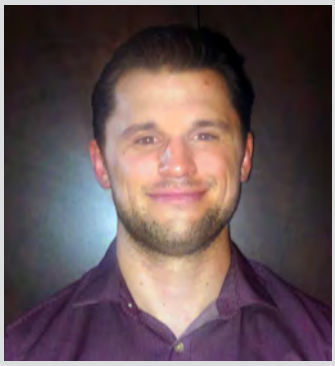
Since 2016, his research has been awarded several prestigious grants, including: a Social Sciences and Humanities Research Council (SSHRC) Insight Development Grant (2016-2019); a Department of National Defence (DND) Defence Engagement Grant (2017); a DND Innovation for Defence Excellence and Security (IDEaS) grant (2018-2019); a Canadian Network for Research on Terrorism, Security, and Society (TSAS) project grant (2018-2019); and a TSAS Major Research Project grant (2019-2021).

His books include *Deterring Rational Fanatics* (University of Pennsylvania Press, 2015) and *Deterring Terrorism: Theory and Practice* (ed., Stanford University Press, 2012), and he has published articles in, among other outlets, *International Security*, *NYU Journal of International Law and Politics*, *Security Studies*, *Journal of Strategic Studies*, *Comparative Strategy*, and *Studies in Conflict and Terrorism*. His commentaries have appeared in the *Globe and Mail*, *National Post*, *Ottawa Citizen*, *Embassy*, *Vanguard*, and *BBC News*.

Prior to joining NPSIA, Professor Wilner held a variety of positions at Policy Horizons Canada – the Government of Canada's strategic foresight laboratory – the Munk School of Global Affairs at the University of Toronto, the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland, and the ETH Zurich in Switzerland.

James Balasch is an MA student at Carleton University's Norman Paterson School of International Affairs. His research focuses on issues surrounding geopolitics, international security, foreign policy and political risk.

Jonathan Kandelshein is an MA student at Carleton University's Norman Paterson School of International Affairs where he focuses on security and defence policy. Apart from his research on the application of artificial intelligence to strategic issues, he also conducts research on foreign disinformation campaigns. He completed his BA at Yeshiva University, graduating summa cum laude in classics and economics, and his JD at Yale Law School. He is currently a summer co-op student for the government in national security.



Cristian Lorenzoni is an MA student at Carleton University's Norman Paterson School of International Affairs. His main interests include intelligence and international affairs along with international political economy. He eventually hopes to pursue a career in political risk analysis.

Sydney Reis is an MA student at Carleton University's Norman Paterson School of International Affairs, specializing in intelligence and international affairs. Her research interests include the intersection of artificial intelligence and ethics, and the role of the Internet in terrorism and foreign interference. Sydney is currently a student policy analyst working with the federal government in national security. She holds a BA in political science from Western University, and has previously worked in various support roles for Ontario Cabinet Ministers.

References

Allen, Greg, and Taniel Chan. 2017. "Artificial Intelligence and National Security." Belfer Center for Science and International Affairs, Harvard Kennedy School, July.

Applebaum, Anne, Edward Lucas, Ben Nimmo, et al. 2018. *Kremlin Disinformation Campaigns: Recommendations to Counter Russia Computational Propaganda in the UK*. London School of Economics. Available at http://www.lse.ac.uk/iga/assets/documents/arena/2018/Recomendations-to-counter-Russia-computational-propaganda-in-the-UK.pdf.

Baines, Paul, and Nicholas O'Shaughnessy. 2013. *Propaganda*, vol. 1. London, England: Sage Library of Military and Strategic Studies.

Benedictus, Leo. 2016. "Invasion of the Troll Armies: From Russian Trump Supporters to Turkish State Stooges." *Guardian*, November 6. Available at https://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian.

Bennett, Lance W., and Steven Livingston. 2018. "The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions." *European Journal of Communication* 33 (2).

Blackwell, Tom. 2019. "How China Uses Shadowy United Front as 'Magic Weapon' to Try to Extend Its Influence in Canada." *National Post*, January 28. Available at https://nationalpost.com/news/how-china-uses-shadowy-united-front-as-magic-weapon-to-try-to-extend-its-influence-in-canada.

Bradshaw, Samantha, and Philip N. Howard. 2018. *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*. Oxford Internet Institute. Available at http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf.

Brattberg, Erik, and Tim Maurer. 2018. *Russian Election Interference: Europe's Counter to Fake News and Cyber Attacks*. Carnegie Endowment For International Peace, May 23. Available at https://carnegieendowment.org/2018/05/23/russian-election-interference-europe-s-counter-to-fake-news-and-cyber-attacks-pub-76435.

Bullock, Joseph, and Miguel Luengo-Oroz. 2019. "Automated Speech Generation from UN General Assembly Statements: Mapping Risks in AI Generated Texts." Paper presented at the International Conference on Machine Learning AI for Social Good Workshop, Long Beach, United States. *arXiv.org*, June 5. Available at https://arxiv.org/pdf/1906.01946.pdf.

Canada. 2015. "Consolidated Federal Laws of Canada, Radio Regulations, 1986." Legislative Services Branch, November 25. Available at https://laws.justice.gc.ca/eng/regulations/SOR-86-982/page-2.html#h-7.

———. 2018a. "Who Said What? The Security Challenges of Modern Disinformation." *Academic Outreach*. Canadian Security Intelligence Service, February.

———. 2018b. "Big Data and the Social Credit System: The Security Consequences." In *China and the Age of Strategic Rivalry*. Canadian Security Intelligence Service, May. Available at https://www.canada.ca/content/dam/csis-scrs/documents/publications/CSIS-Academic-Outreach-China-report-May-2018-en.pdf.

———. 2018c. "Canada's Democratic Process." Canadian Centre for Cyber Security, August 15. Available at https://cyber.gc.ca/en/.

———. 2018d. "Consolidated Federal Laws of Canada, Television Broadcasting Regulations, 1987." Legislative Services Branch, September 1. Available at https://laws.justice.gc.ca/eng/regulations/SOR-87-49/page-2.html#h-5.

———. 2019. "G7 Rapid Response Mechanism." Government of Canada, January 30. Available at https://www.canada.ca/en/democratic-institutions/news/2019/01/g7-rapid-response-mechanism.html.

Carrell, Severin. 2017. "Russian Cyber-activists 'Tried to Discredit Scottish Independence Vote'." *Guardian*, December 13. Available at https://www.theguardian.com/politics/2017/dec/13/russian-cyber-activists-tried-to-discredit-scottish-independence-vote-says-analyst.

Chesney, Robert, and Danielle K. Citron. 2018. *Disinformation on Steroids: The Threat of Deep Fakes*. Council on Foreign Relations, October 16. Available at https://www.cfr.org/report/deep-fake-disinformation-steroids.

Chhor, Khatya. 2017. "As French Media Went Dark, Bots and Far-right Activists Drove #MacronLeaks." *France 24*, May 9. Available at https://www.france24.com/en/20170508-french-media-blackout-bots-far-right-activists-wikileaks-pushed-macronleaks.

Chien, Li-chung, Chung Li-hua, and Jonathan Chin. 2018. "China Using Fake News to Divide Taiwan." *Taipei Times*, September 18. Available at http://www.taipeitimes.com/News/front/archives/2018/09/16/2003700513.

Cleary, Gillian. 2019. "Twitterbots: Anatomy of a Propaganda Campaign." *Symantec Blogs*, June 5. Available at https://www.symantec.com/blogs/threat-intelligence/twitterbots-propaganda-disinformation.

Cole, J. Michael. 2018. "The Impact of China's Disinformation Operations against Taiwan." Prospect Foundation, October 3. Available at https://www.pf.org.tw/article-pfch-2049-6365.

Counts, Laura. 2019. "How Information Is Like Snacks, Money, and Drugs—to Your Brain." *Haas News*, June 19. Available at https://newsroom.haas.berkeley.edu/how-information-is-like-snacks-money-and-drugs-to-your-brain/.

CREST (Centre for Research and Evidence on Security Threats). 2017. "Russian Influence and Interference Measures Following the 2017 UK Terrorist Attacks." *Policy Brief.* Cardiff University Crime and Security Research Institute. Available at https://www.cardiff.ac.uk/news/view/1037714-russian-influence-and-interference-measures-following-the-2017-uk-terrorist-attacks.

deepfakeapp. 2018. "How to Use FakeApp to Make Your Own Faceswap Gifs." *Reddit*, February 7. Available at https://www.reddit.com/r/GifFakes/comments/7w1pp9/how_to_use_fakeapp_to_make_your_own_faceswap_gifs/.

Dem. 2019. "When Russia Uses 'Deep Fakes' against Our Nominee, We Need to Be Prepared." *Daily Kos* (Community), June 16. Available at https://www.dailykos.com/stories/2019/6/16/1865186/-When-Russia-uses-Deep-Fakes-against-our-nominee-we-need-to-be-prepared.

DFRLab. 2017a. "Hashtag Campaign: #MacronLeaks." *Medium*, May 5. Available at https://medium.com/dfrlab/hashtag-campaign-macronleaks-4a3fb870c4e8.

———. 2017b. "#ElectionWatch: Scottish Vote, Pro-Kremlin Trolls." *Medium*, December 13. Available at https://medium.com/dfrlab/electionwatch-scottish-vote-pro-kremlin-trolls-f3cca45045bb.

———. 2019. "Top Takes: Suspected Russian Intelligence Operation." *Medium*, June 22. Available at https://medium.com/dfrlab/top-takes-suspected-russian-intelligence-operation-39212367d2f0.

DiResta, Renee, John Little, Jonathon Morgan, Lisa Maria Neudert, and Ben Nimmo. 2017. "The Bots That Are Changing Politics." *Vice*, November 2. Available at: https://motherboard.vice.com/en_us/article/mb37k4/twitter-facebook-google-bots-misinformation-changing-politics.

DiResta, Renee, Kris Shaffer, Becky Ruppel, et al. 2018. *The Tactics & Tropes of the Internet Research Agency.* New Knowledge, December. Available at https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper.pdf

Dittrich, Paul-Jasper. 2018. *Better Together? Franco-German Cooperation on AI*. Berlin: Jacque Delors Institute, December 18.

Dixson, Mary. 2005. "A Review of 'Civic Literacy: How Informed Citizens Make Democracy Work'." *Political Communication* 22 (2): 245.

EU. 2018. "EU Member States Sign up to Cooperate on Artificial Intelligence." European Commission, April 10. Available at https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence.

Fleishman, Glenn. 2019. "How to Spot the Realistic Fake People Creeping into Your Timelines." *Fast Company*, April 30. Available at https://www.fastcompany.com/90332538/how-to-spot-the-creepy-fake-faces-who-may-be-lurking-in-your-timelines-deepfaces.

Guardian. 2019. "Real v Fake: Debunking the 'Drunk' Nancy Pelosi Footage – Video." *Guardian*, May 24. Available at https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video.

Gunter, Joel, and Olga Robinson. 2018. "Sergei Skripal and the Russian Disinformation Game." *BBC*, September 9. Available at https://www.bbc.com/news/world-europe-45454142.

Harwell, Drew. 2019a. "Faked Pelosi Videos, Slowed to Make Her Appear Drunk, Spread across Social Media." *Washington Post*, May 24. Available at https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media.

———. 2019b. "Facebook Acknowledges Pelosi Video Is Faked but Declines to Delete It." *Washington Post*, May 24. Available at https://www.washingtonpost.com/technology/2019/05/24/facebook-acknowledges-pelosi-video-is-faked-declines-delete-it/.

Horton, Chris. 2018. "Specter of Meddling by Beijing Looms over Taiwan's Elections." *New York Times*, November 22. Available at https://www.nytimes.com/2018/11/22/world/asia/taiwan-elections-meddling.html.

Howard, Philip N., Bharath Ganesh, Dimitra Liotsiou, John Kelly, and Camille François. 2018. *The IRA, Social Media and Political Polarization in the United States, 2012-2018*. University of Oxford.

Huggler, Justin. 2016. "German Teenager 'Made up' Migrant Rape Claim." *Telegraph*, January 31. Available at https://www.telegraph.co.uk/news/worldnews/europe/germany/12132687/German-teenager-made-up-migrant-rape-claim.html.

Jianguo123. 2018. "Deepfakes Creator for Hire." *Reddit*, May 25. Available at https://www.reddit.com/r/SFWdeepfakes/comments/8m63cn/deepfakes_creator_for_hire/.

Kim, Mie (ed.). 2018. *Closing the Digital Loopholes that Pave the Way for Foreign Interference in U.S. Elections*. CLC, April 16.

King, Esther. 2016. "Russian Hackers Targeting Germany: Intelligence Chief." *Politico*, November 29. Available at https://www.politico.eu/article/german-intelligence-chief-russian-hackers-targeting-us-bruno-kahl-vladimir-putin/.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Medial Posts, for Strategic Distraction, not Engaged Argument." *American Political Science Review* 111 (3).

Knight, Will. 2018. "The Defense Department Has Produced the First Tools for Catching Deep Fakes." *MIT Technology Review*, August 7. Available at https://www.technologyreview.com/s/611726/the-defense-department-has-produced-the-first-tools-for-catching-deepfakes/.

Ko, Tin-yau. 2018. "How Fake News Led to Suicide of Taiwan Representative in Osaka." *ejinsight*, September 19. Available at http://www.ejinsight.com/20180919-how-fake-news-led-to-suicide-of-taiwan-representative-in-osaka/.

Lee, Kai Fu. 2018. "Why China Can Do AI More Quickly and Effectively than the US." *Wired*, October 23. Available at https://www.wired.com/story/why-china-can-do-ai-more-quickly-and-effectively-than-the-us/.

Leslie, Tim, Nathan Hoad, and Ben Spraggon. 2018. "How Hard Is It to Make a Believable Deepfake?" *ABC News (Australia)*, October 2. Available at https://www.abc.net.au/news/2018-09-28/fake-news-how-hard-is-it-to-make-a-deepfake-video/10313906.

Liwaiwai. 2019. "How Artificial Intelligence Can Detect – and Create – Fake News." *liwaiwai.com*, July 3. https://liwaiwai.com/2019/07/03/how-artificial-intelligence-can-detect-and-create-fake-news/.

Lowe, Ryan. 2019. "OpenAI's GPT-2: The Model, the Hype, and the Controversy." *Medium*, February 18. Available at https://towardsdatascience.com/openais-gpt-2-the-model-the-hype-and-the-controversy-1109f4bfd5e8.

Mack, David. 2018. "This PSA about Fake News from Barack Obama Is Not What It Appears." *BuzzFeed*, April 17. Available at https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed#.gcxNolpGL.

Mak, Aaron. 2019. "When Is Technology Too Dangerous to Release to the Public?" *Slate*, February 22. Available at https://slate.com/technology/2019/02/openai-gpt2-text-generating-algorithm-ai-dangerous.html.

Mahmood, Asad. 2019. "A Look into GLTR (Using GPT-2)." *Medium*, April 21. Available at https://towardsdatascience.com/a-look-into-gltr-using-gpt-2-76d823057421.

Martin, Michelle. 2019. "German Spy Agency Probes Russia Links to Right-wing Parties – Report." *Reuters*, February 14. Available at https://uk.reuters.com/article/uk-germany-security/german-spy-agency-probes-russia-links-to-right-wing-parties-report-idUKKCN1Q31MD.

Maynes, Charles. 2018. "A Mole among Trolls: Inside Russia's Online Propaganda Machine." *Public Radio International*, March 16. Available at https://www.pri.org/stories/2018-03-16/mole-among-trolls-inside-russias-online-propaganda-machine.

Meister, Stefan. 2016. "The 'Lisa Case': Germany As a Target of Russian Disinformation." *NATO Review*. Available at https://www.nato.int/DOCU/review/2016/Also-in-2016/lisa-case-germany-target-russian-disinformation/EN/index.htm.

Mohan, Meghan. 2017. "Macron Leaks: The Anatomy of a Hack." *BBC*, May 9. Available at https://www.bbc.com/news/blogs-trending-39845105.

Monaco, Nicholas J. 2017. "Computational Propaganda in Taiwan: Where Digital Democracy Meets Automated Autocracy." In *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, edited by Samuel Woolley and Philip N. Howard. Oxford, UK: Oxford University Press. Available at http://comprop.oii.ox.ac.uk/.

Mozur, Paul. 2018. "A Genocide Incited on Facebook, with Posts from Myanmar's Military." *New York Times*, October 15, sec. Technology. Available at https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

Musil, Steven. 2019. "Videos Doctored to Make Pelosi Appear Drunk Spread across Social Media." *CNET*, May 24. Available at https://www.cnet.com/news/videos-doctored-to-make-pelosi-appear-drunk-spread-across-social-media/.

Nieva, Richard. 2018. "Alphabet Chairman Says Google Duplex Passes Turing Test in One Specific Way." *CNET*, May 10. Available at https://www.cnet.com/news/alphabet-chairman-says-google-duplex-passes-turing-test-in-one-specific-way-io-2018/.

Nimmo, Ben. 2016. *Sputnik: Propaganda in a New Orbit*. CEPA, January. Available at https://web.archive.org/web/20170722173531/http://cepa.org/files/?id_plik=2083.

NowThis Politics. 2019. "Fake, Doctored Footage of a 'Sickly' Nancy Pelosi Spreads on Fox News and Social Media." *Facebook*, May 28. Available at https://www.facebook.com/watch/?v=2318188621606677.

Piccone, Ted. 2018. "Democracy and Digital Technology." *Sur International Journal on Human Rights* 15 (27): 29.

Piper, Kelsey. 2019. "These Fake Images Tell a Scary Story of How Far AI Has Come." *Vox*, May 31. Available at https://www.vox.com/future-perfect/2019/5/31/18645993/ai-deepfakes-gan-explained-machine-learning.

Polyakova, Alina. 2018. *Weapons of the Weak: Russia and AI-driven Asymmetric Warfare*. Brookings, November 15. Available at https://www.brookings.edu/research/weapons-of-the-weak-russia-and-ai-driven-asymmetric-warfare/.

Rocha, Roberto. 2018. "Data Sheds Light on How Russian Twitter Trolls Targeted Canadians." *CBC News*, August 3. Available at https://www.cbc.ca/news/canada/russian-twitter-trolls-canada-targeted-1.4772397.

Silverman, Craig. 2018. "How to Spot a Deepfake Like the Barack Obama–Jordan Peele Video." *BuzzFeed*, April 17. Available at https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed?utm_term=.bxbj7Rqm7#.hsZ0VPeyV.

Solsman, Joan E. 2019. "Deepfakes May Ruin the World. And They Can Come for You, Too." *CNET*, April 4. Available at https://www.cnet.com/news/deepfakes-may-try-to-ruin-the-world-but-they-can-come-for-you-too/.

Sputnik. 2017. "Ex-French Economy Minister Macron Could Be 'US Agent' Lobbying Banks' Interests." *Sputnik News*, February 4. Available at https://sputniknews.com/analysis/201702041050340451-macron-us-agent-dhuicq/.

Stelzenmuller, Constanze. 2017. "Testimony – The impact of Russian interference on German's 2017 elections." Brookings Institute, June 28. Available at https://www.brookings.edu/testimonies/the-impact-of-russian-interference-on-germanys-2017-elections/.

Tillman, Maggie. 2019. "Deepfake Videos Explained: What Are They and Should You Be Worried?" *Pocket-lint*, June 14. Available at https://www.pocket-lint.com/apps/news/148360-deepfake-videos-explained-what-are-they-and-should-you-be-worried.

Toucas, Boris. 2017. "The Macron Leaks: The Defeat of Informational Warfare." CSIS, May 30. Available at https://www.csis.org/analysis/macron-leaks-defeat-informational-warfare.

Tsai, Michelle. 2018. "The China Factor in Taiwan's Local Elections." *The Diplomat*, December 6. Available at https://thediplomat.com/2018/12/the-china-factor-in-taiwans-local-elections/.

Tung, Wan Qing. 2016. "Have You Get Shocked Today?! How Content Farms Generate Affective Publics in Cyberspace..." *Cultural Studies@Lingnan* 51 (1). Available at http://commons.ln.edu.hk/mcsln/vol51/iss1/7/.

UK. 2018. *Disinformation and 'Fake News': Interim Report: Government Response to the Committee's Fifth Report of Session 2017–19*. House of Commons Digital, Culture, Media and Sport Committee, UK Parliament, October 17. Available at https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1630/1630.pdf.

———. 2019. "Foreign Influence in Political Campaigns." House of Commons Digital, Culture, Media and Sport Committee, UK Parliament, February 18, 69-71. Available at https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/179109.htm.

United Nations. n.d. "OHCHR Rule of Law – Democracy and Human Rights." United Nations Human Rights, Office of the High Commissioner. Available at https://www.ohchr.org/en/Issues/RuleOfLaw/Pages/Democracy.aspx.

Villasenor, John. 2019. "Artificial Intelligence, Deepfakes, and the Uncertain Future of Truth." Brookings, February 14. Available at https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/.

Vilmer, Jean-Baptiste Jeangène. 2019. *The "Macron Leaks" Operation: A Post-Mortem*. Atlantic Council, June. Available at https://www.atlanticcouncil.org/images/publications/The_Macron_Leaks_Operation-A_Post-Mortem.pdf.

Vilmer, Jean-Baptiste Jeangène, Alexandre Escorcia, Marine Guillaume, and Janaina Herrera. 2018. "Information Manipulation: A Challenge for Our Democracies." Policy Planning Staff (CAPS) of the Ministry for Europe and Foreign Affairs and the Institute for Strategic Research (IRSEM) of the Ministry for the Armed Forces, Paris, August.

Volz, Dustin. 2017. "U.S. Far-right Activists, WikiLeaks and Bots Help Amplify Macron Leaks: Researchers." *Reuters*, May 6. Available at https://www.reuters.com/article/us-france-election-cyber-idUSKBN1820QO.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380). Available at https://science.sciencemag.org/content/359/6380/1146.

Xiao, Liu, Bin Zhang, Anjana Susarla, and Rema Padman. 2019. "Go to YouTube and Call Me in the Morning: Use of Social Media for Chronic Conditions." *MIS Quarterly*. Available at https://ssrn.com/abstract=3061149.

Xiao, Qiang. 2011."Leaked Propaganda Directives and Banned 'Future'." *China Digital Times*, June 24. Available at https://chinadigitaltimes.net/2011/06/future-banned-on-sina-weibo-search/.

Zakharov, Egor. 2019. "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models." *YouTube*, May 21. Available at https://www.youtube.com/watch?time_continue=2&v=p1b5aiTrGzY.

Zakharov, Egor, Aliaksandra Shysheya, Egor Burkov, and Victor Lempitsky. 2019. "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models." *arXiv.org*, May 20. Available at https://arxiv.org/pdf/1905.08233v1.pdf.

Zimmer, Nathaniel, and Bob Erskine. 2018. "Canada Works with Allies to Combat Foreign Election Interference." *Toronto Star*, July 20. Available at http://search.proquest.com/docview/2072853037/citation/FAE50F924B0F45F1PQ/1.

Zucconi, Alan. 2018. "How to Create the Perfect DeepFakes." *alanzucconi.com*, March 14. Available at https://www.alanzucconi.com/2018/03/14/create-perfect-deepfakes/.

Endnotes

1 According to a recent Symantec report, the IRA's 3836 fake accounts (with 123 main accounts and 3713 auxiliary accounts) pushed out over 10 million tweets (Cleary 2019).

2 Of course, this does not mean that Russia does not favour certain politicians or political parties. For example, the German domestic intelligence agency (*Bundesamt für Verfassungsschutz*) has been investigating Russia's possible ties to far-right parties, including the Alternative for Germany party (Martin 2019).

3 The Turing Test "is a way of evaluating a machine's intelligence – to pass, a robot must behave in a way indistinguishable to a human" (Nieva 2018).

4 DARPA is the Defense Advanced Research Projects Agency.

MLI

True North in Canadian public policy

Critically Acclaimed, Award-Winning Institute

The Macdonald-Laurier Institute fills a gap in Canada's democratic infrastructure by focusing our work on the full range of issues that fall under Ottawa's jurisdiction.

- One of the top five think tanks in Canada and No. 1 in Ottawa according to the University of Pennsylvania.
- Cited by five present and former Canadian Prime Ministers, as well as by David Cameron, the British Prime Minister.
- First book, *The Canadian Century: Moving out of America's Shadow*, won the Sir Antony Fisher International Memorial Award in 2011.
- *Hill Times* says Brian Lee Crowley is one of the 100 most influential people in Ottawa.
- The *Wall Street Journal*, the *Economist*, the *Globe and Mail*, the *National Post* and many other leading national and international publications have quoted the Institute's work.

"The study by Brian Lee Crowley and Ken Coates is a 'home run'. The analysis by Douglas Bland will make many uncomfortable but it is a wake up call that must be read."
former Canadian Prime Minister Paul Martin on MLI's project on Aboriginal people and the natural resource economy.

Ideas Change the World

Independent and non-partisan, the Macdonald-Laurier Institute is increasingly recognized as the thought leader on national issues in Canada, prodding governments, opinion leaders and the general public to accept nothing but the very best public policy solutions for the challenges Canada faces.

Where You've Seen Us

CTV

cpac FOR THE RECORD

corus ENTERTAINMENT

CBC news

FP Foreign Policy

tvo

The Economist

WALL STREET JOURNAL

THE HILL TIMES

THE GLOBE AND MAIL
CANADA'S NATIONAL NEWSPAPER • FOUNDED 1844

NATIONAL POST

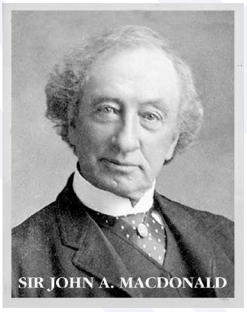
About the Macdonald-Laurier Institute

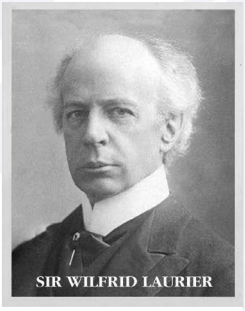
What Do We Do?

When you change how people think, you change what they want and how they act. That is why thought leadership is essential in every field. At MLI, we strip away the complexity that makes policy issues unintelligible and present them in a way that leads to action, to better quality policy decisions, to more effective government, and to a more focused pursuit of the national interest of all Canadians. MLI is the only non-partisan, independent national public policy think tank based in Ottawa that focuses on the full range of issues that fall under the jurisdiction of the federal government.

What Is in a Name?

The Macdonald-Laurier Institute exists not merely to burnish the splendid legacy of two towering figures in Canadian history – Sir John A. Macdonald and Sir Wilfrid Laurier – but to renew that legacy. A Tory and a Grit, an English speaker and a French speaker – these two men represent the very best of Canada's fine political tradition. As prime minister, each championed the values that led to Canada assuming her place as one of the world's leading democracies. We will continue to vigorously uphold these values, the cornerstones of our nation.


SIR JOHN A. MACDONALD


SIR WILFRID LAURIER

Working for a Better Canada

Good policy doesn't just happen; it requires good ideas, hard work, and being in the right place at the right time. In other words, it requires MLI. We pride ourselves on independence, and accept no funding from the government for our research. If you value our work and if you believe in the possibility of a better Canada, consider making a tax-deductible donation. The Macdonald-Laurier Institute is a registered charity.

Our Issues

The Institute undertakes an impressive program of thought leadership on public policy. Some of the issues we have tackled recently include:

- Aboriginal people and the management of our natural resources;
- Making Canada's justice system more fair and efficient;
- Defending Canada's innovators and creators;
- Controlling government debt at all levels;
- Advancing Canada's interests abroad;
- Ottawa's regulation of foreign investment; and
- How to fix Canadian health care.

True North in
Canadian public policy

CONTACT US: Macdonald-Laurier Institute
323 Chapel Street, Suite #300
Ottawa, Ontario, Canada
K1N 7Z2

TELEPHONE: (613) 482-8327

WEBSITE: www.MacdonaldLaurier.ca

CONNECT WITH US:

@MLInstitute

www.facebook.com/
MacdonaldLaurierInstitute

www.youtube.com/
MLInstitute

What people are saying about the Macdonald-Laurier Institute

In five short years, the institute has established itself as a steady source of high-quality research and thoughtful policy analysis here in our nation's capital. Inspired by Canada's deep-rooted intellectual tradition of ordered liberty – as exemplified by Macdonald and Laurier – the institute is making unique contributions to federal public policy and discourse. Please accept my best wishes for a memorable anniversary celebration and continued success.

THE RIGHT HONOURABLE STEPHEN HARPER

The Macdonald-Laurier Institute is an important source of fact and opinion for so many, including me. Everything they tackle is accomplished in great depth and furthers the public policy debate in Canada. Happy Anniversary, this is but the beginning.

THE RIGHT HONOURABLE PAUL MARTIN

In its mere five years of existence, the Macdonald-Laurier Institute, under the erudite Brian Lee Crowley's vibrant leadership, has, through its various publications and public events, forged a reputation for brilliance and originality in areas of vital concern to Canadians: from all aspects of the economy to health care reform, aboriginal affairs, justice, and national security.

BARBARA KAY, NATIONAL POST COLUMNIST

Intelligent and informed debate contributes to a stronger, healthier and more competitive Canadian society. In five short years the Macdonald-Laurier Institute has emerged as a significant and respected voice in the shaping of public policy. On a wide range of issues important to our country's future, Brian Lee Crowley and his team are making a difference.

JOHN MANLEY, CEO COUNCIL