



The Economics of Deepfakes

Nir Kshetri^{ID}, The University of North Carolina at Greensboro

This article explores the diverse motivations and the cost–benefit structure of nefarious actors engaged in creating and distributing deepfakes.

Extrinsic as well as intrinsic motivations are analyzed.

The creation, distribution, and use of deepfakes have skyrocketed in recent years. While deepfake videos began with the invention of the Video Rewrite program in 1997, the release of deepfake tools such as DeepFaceLab in 2018 accelerated deepfake production.¹ By early 2023, the number of deepfakes online was estimated to be in the millions (Figure 1).² A study of Sensity artificial intelligence (AI), which specializes in deepfake content, found that 90–95% of deepfake videos published online during 2018 to 2020 were primarily based on nonconsensual pornography.³ Deviant behaviors such as this can be viewed as a combination of two offenses: deviant online behavior and sexual offense. Moreover, many deepfake audios, videos, and images used in scams are not publicly available.

With the advancement in deepfake creation tools, perpetrators are likely to use such tools to commit

more serious cybercrimes. For instance, cybercriminals may hijack Internet-of-Things devices that use voice or face recognition, such as Amazon's Alexa.⁴

There are significant economic and social costs associated with deepfakes. For instance, the maximum reported loss from a single cybercrime incident involving deepfake was about US\$35 million. On the social front, deepfakes are viewed as real threats to women and girls. This is because most of the current deepfakes involve pornographic videos and images that target and abuse women and girls, who have no way of preventing an offender from creating such content. Some view deepfakes as a new form of gender-based violence, which is being used to exploit, humiliate, and harass women.⁵

Deepfakes may also pose a threat to democracy. AI has arguably already developed at a level that can pass off as a politician without being detected easily. It is even likely that the effect could be large enough to change the outcome of a future presidential election in the United States.⁶

An attack involving deepfakes can also cause an organization reputation loss and may lead to other costs. For instance, an ill-motivated executive would abuse deepfakes and direct attacks against an adversary company.

Digital Object Identifier 10.1109/MC.2023.3276068
Date of current version: 26 July 2023

A deepfake video of a CEO saying that their company will not meet targets could lead to a significant decline in share price.⁷ Using AI-generated profile photos and AI-written posts, a fake account could earn many followers. A large network of such accounts can be used to engage in actions that can damage a company's reputation.⁸

In this article, we examine the diverse motivations of actors engaged in creating and distributing deepfakes. We also give an overview of the cost-benefit structure of such actors to engage in deepfake-related offenses.

A TYPOLOGY OF DEEPFAKES

Table 1 presents a classification and examples of major types of deepfakes. We first discuss the two axes in Table 1.

Motivation (the horizontal axis)

The creation and distribution of deepfakes may involve extrinsic as well as intrinsic motivation. Economics theory suggests that human behavior is a result of "incentives applied from outside the person."¹¹ Externally motivated deepfake offenders have thus an interest in revenue-generating deepfake activities with high financial returns.

The idea behind intrinsic motivation is that human need for competence and self-determination are linked with interest and enjoyment.¹² Intrinsically motivated individuals do activities for "inherent satisfactions rather than for some separable consequence."¹³ An intrinsically motivated person acts "for the fun or challenge entailed rather than because of external prods, pressures or rewards."

User environment (the vertical axis)

While some deepfake apps are customized for a single user, others engage a large number of users. The former type is often

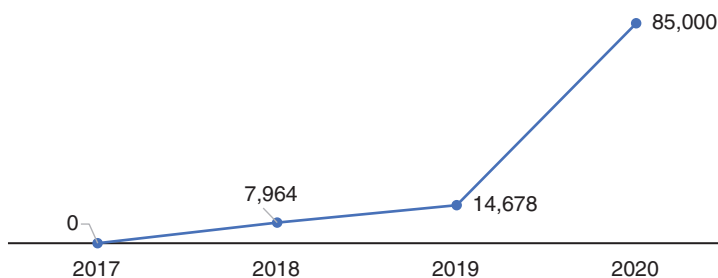


FIGURE 1. The number of deepfake videos available online. Data source: 2018 (December) and 2019 (October) Deeptrace Labs.⁹ For 2017 and 2020 (December), Sensity.¹⁰

used by resourceful actors, such as government agencies or experienced cybercriminals to launch elaborate and targeted cyberattacks. The latter type is created for financial gains as well as for nonfinancial motivations, such as fun, and aims at attracting a large number of users.

The four cells in Table 1

Cell I. Deepfakes or apps to create them are offered for users with an emphasis on monetizing the user base. An October 2020 report by Sensity provides a glimpse into a well-developed deepfake ecosystem on the messaging platform Telegram, which provides clear evidence that there is likely a substantial monetary benefit in the deepfake ecosystem. An AI-powered bot allowed users to create photo-realistically "strip naked" images of women. The bot provided a free and simple user interface that functioned on smartphones and computers, which dramatically increases accessibility and ease of use. Users receive processed stripped images simply by uploading the targets' picture to the bot following a short generation process. Sensity found personal "stripped" images of at least 104,852 women, which were shared publicly. About 70% of the targets were private individuals. The bot

TABLE 1. Diverse types of deepfakes: A typology and examples.

Motivation User environment	Extrinsic	Intrinsic
Created for a multiuser environment	Cell I Deepfakes or apps to create them offered for users with an emphasis on monetizing [for example, a bot to create "stripped" images shared via VK (generated substantial advertising revenues)]	Cell II Deepfake apps created for hobby and shared in a community [for example, deepfakes subreddit (SFW)]
Customized for a single user	Cell III Elaborate and targeted financially motivated social engineering schemes (examples in Table 2)	Cell IV Created by companies and governments to improve their reputation or damage the reputations of their adversaries (e.g., Spamouflage).

and affiliated channels had more than 100,000 members worldwide, 70% of whom were from Russia and former Soviet Union countries. The bot generated substantial advertising revenues via the Russian online social media and social networking service VK. VK featured related activity in 380 pages.¹⁴ Many of the victims were underage girls.¹⁵ According to *Wired*, in October 2020 the bot sent out a gallery of new images to an associated Telegram channel which has almost 25,000 subscribers. Most images were viewed more than 3,000 times. A separate Telegram channel promoting the bot had over 50,000 subscribers.¹⁶

Cell II. Intrinsic motivation can be 1) enjoyment-based and 2) obligation/community-based. We illustrate these ideas with examples from social news aggregation, content rating, and the discussion website Reddit. In the deepfakes subreddit (SFW) (<https://www.reddit.com/r/SFWdeepfakes/>), people create deepfakes for a hobby rather than for money.¹⁷ This can be viewed as an enjoyment-based intrinsic motivation.

Some individuals also derive enjoyment by engaging in a challenging task, or becoming immersed in the process, which is also an intrinsic motivation.¹⁸ The creator of viral Tom Cruise deepfake videos put the issue this way: “My talent is my eye for perfection and my high-quality standards. I’d rather finish something I’m proud of than quick garbage.”¹⁹

We return to the Reddit example to illustrate obligation/community-based

intrinsic motivation. A Reddit user’s karma, which is publicly displayed, reflects how much their contributions mean to the community. To increase karma points, a Reddit user can submit links (for example, of deepfakes) and comments, which is an obligation/community-based intrinsic motivation.

Cell III. Deepfakes are being used in elaborate and targeted financially motivated social engineering schemes. Some such examples are presented in Table 2. We briefly discuss them here:

- › In 2019, a Californian woman was scammed out of US\$300,000 using deepfakes. The scheme involved two different accounts used by the same or two scammers. The first criminal on an online dating site pretended to be Sean Buck, the U.S. Navy vice admiral and the superintendent of the Naval Academy. The victim talked to the fake Buck on Skype regularly. The second scammer pretended to be an American and after months of communication told the victim that he was in a foreign prison. The victim asked the fake vice admiral to help to release the prisoner. The fake Buck asked the victim to transfer money “to pay the lawyer,” law enforcement agencies discovered, when the victim watched deepfake clips of Buck speaking.²⁰
- › Deepfake voice phishing (vishing)⁸ is becoming common.

This technique involves using a cloned voice to exploit the victim’s professional or personal relationships. In 2019, cybercriminals used a deepfake to fool the CEO of a United Kingdom-based energy firm into making a US\$234,000 wire transfer. The CEO thought that the person speaking on the phone was the chief executive of the firm’s German parent company. The deepfake caller asked him to transfer funds to a Hungarian supplier within an hour, emphasizing that the matter was extremely urgent. The fraudsters used AI-based software to successfully imitate the German executive’s voice.²¹

- › Cybercriminals are exploiting social media networks with deepfakes. Japanese Manga artist Chikae Ide was scammed out of more than US\$500,000 using deepfakes. The scammer created a deepfake version of American actor and producer Mark Ruffalo to engage in a fake romance scam and contacted Ide on Facebook. The scammer built an emotional trust with Ide and continually asked her for thousands of dollars, saying that they needed money for a plane ticket, hospital bills, and other things. The criminal also faked having cancer to get money.²² A friend warned Ide that the online romance partner wrote “like somebody who has not learned English” and thus couldn’t be Ruffalo.

TABLE 2. Some high-profile deepfake attacks launched for financial gains.

Year	Victim and location	Type of attack	Amount scammed
2019	A Californian woman	Deepfake videos	About US\$300,000
2019	United Kingdom-based energy company	Vishing	US\$234,000
2020	A United Arab Emirates bank	Deepfake voice technology	>US\$35 million
2018–2021	Japanese Manga artist	Deepfake videos	>US\$500,000

But a 30-second deepfake video call falsely convinced Ide that the person behind the screen chatting with her was real Ruffalo.²³

- › In a high-profile case, in January 2020, cybercriminals defrauded a United Arab Emirates bank of over US\$35 million using a deepfake voice technology. The technology was used to imitate a company director, who was known to a bank manager. The manager authorized the transactions.²⁴

Cell IV. Deepfakes created by government agencies to improve their reputation or damage the reputations of their adversaries belong to this category. For example, since late 2022, a Chinese Communist Party-aligned influence operation, known as *Spamouflage*, has been allegedly using deepfake news anchors to promote China's global role and spread disinformation against the United States. The social media analytics firm Graphika has described two videos circulating on social media platforms that featured a male and a female anchor. Both anchors are Caucasian-looking and speak English. The videos used the logo of a media company called Wolf News, which is believed to be fictitious. The company's accompanying slogan is "Focus on hot spots and broadcast in real time." The male anchor criticized that, in the context of gun violence, the U.S. government has engaged in a "hypocritical repetition of empty rhetoric." The female anchor emphasized the importance of China–United States cooperation for the recovery of a global economy.²⁵ This is the first known example of deepfake technology's use for state-aligned influence operations.²⁶

COST–BENEFIT ANALYSIS OF ENGAGING IN DEEPFAKE OFFENSES

An offender engages in deepfake-related activities if²³

$$M_b + P_b > I_c + O_{1c} + P_c + O_{2c} \pi_{arr} \pi_{con} \quad (1)$$

where:

- › M_b = monetary benefits from the creation/distribution of deepfakes
- › P_b = psychological benefits from the creation/distribution of deepfakes
- › I_c = direct investment costs
- › O_{1c} = opportunity costs to engage in the creation/distribution of deepfakes
- › P_c = psychological costs of the engagement in the creation/distribution of deepfakes
- › O_{2c} = monetary opportunity costs of conviction
- › π_{arr} = probability of arrest
- › π_{con} = probability of conviction.

The term $O_{2c} \pi_{arr} \pi_{con}$ is referred to as the *expected penalty effect*.

The benefits

M_b . From a financially motivated cybercriminal's standpoint, there are substantial monetary incentives for engaging in cybercrimes using deepfakes. Table 2 provides four such examples.

Many other cybercrimes have been reported. For instance, in India, the Uttar Pradesh state Police's cybercrime cell reported over 200 such complaints in early 2021, in which pornographic videos were used to extort victims. The ransom ranged from about US\$61 to more than US\$610. The gangs threatened victims that their pornographic video would be posted on social media. The criminals used WhatsApp, Facebook, and Instagram to collect contact details and pictures of victims, mainly businessmen, professionals, and students. They also created fake social media profiles of females. The perpetrators contacted the victims and used prerecorded videos of females to engage in conversation. The frames from the videos were superimposed on pornographic clips.²⁷

Deepfakes are also used in content monetization in social media. Convincing deepfakes can quickly reach millions of people.²⁸ Content monetization in social media requires maximizing

engagement. The quality of the content is irrelevant. Social media companies try to keep people on their platforms longer to deliver ads.²⁹

P_b . Psychological benefits are related to intrinsic motivation. As discussed above, intrinsically motivated individuals engage in activities for satisfaction. For instance, psychopaths are often happy when others suffer.

The use of deepfakes by state-aligned influence operations (for example, Spamouflage's use of deepfake news anchors) also falls in this category. State-sponsored disinformation campaigns and cyberwars are often held and fought for intangible goals, such as dominance and prestige. Some view deepfake videos as part of a larger misinformation and disinformation ecosystem.²⁹ Deepfakes are being used by enemy and adversary states to spread misinformation and disinformation and pursue their national interests. For instance, in the Russo–Ukrainian war, both Russia and Ukraine have deployed deepfake videos against each other.³⁰

The costs

I_c . AI in creating deepfakes is becoming easy to use, resulting in low investment costs. No sophisticated knowledge is required to generate deepfakes. For instance, the AI-powered lip sync app Wombo AI, which is trained with video recordings of real performers, makes it possible to convert any person's photo into a video clip with just a few clicks. The person will be singing well-known songs with matching facial expressions.¹

Individuals at the marketplace Fiverr.com can create deepfakes for as little as US\$5.¹⁷ Some charge just US\$20 to create fabricated videos of exes, coworkers, friends, enemies, and classmates.³¹

O_{1c} . Opportunity costs vary depending on the availability of jobs. An increase in unemployment rate decreases the opportunity cost of crime. As noted above, deepfake-related cybercrimes are growing in India. More broadly, an increasing number

of young people have been reported to engage in cybercrimes due to improved Internet connectivity and increased unemployment. The Covid-19 pandemic further worsened the situation.³²

P_c. Psychological costs are associated with the psychological and mental energy needed in creating and distributing deepfakes (for example, fear of punishment, guilt). The feeling of guilt is not equally pervasive across all offenders. The level of guilt and shame is also a function of demographic, psychographic, and social characteristics of offenders and victims.


Regarding the characteristics of offenders, a study found that psychopathic personality traits were positively related to the creation and dissemination of deepfake pornography. The study's participants scoring higher on a measure of psychopathy were less likely to view the situation as harmful. They were also less likely to believe that it was a crime to engage in such acts. They were more likely to believe that the victim was to blame for the incident. Individuals with a higher degree of psychopathy are likely to have a higher propensity to engage in online offenses involving the creation and sharing of deepfake pornographic images. The researchers also found that deepfake pornography depicting female victims was associated with greater perceptions of harm and criminality than cases with male victims.

O_{2c}. While some jurisdictions have laws against some forms of deepfakes, most countries haven't enacted any laws at all against deepfakes. For instance, In the United States, while 46 states have some ban on revenge porn, only Virginia and California include faked and deepfaked media. Canada and the United Kingdom reported to have regulatory and legal voids in the areas of deepfaked pornographic material.³³ In the United Kingdom, revenge porn is banned, but the law fails to protect deepfake victims. No other country bans fake nonconsensual porn at a

national level.¹⁵ This means that the opportunity cost of conviction for most deepfake-related offenses in most jurisdictions is zero.

π_{arr} and π_{con} . It is difficult to investigate deepfake crimes and prosecute and convict offenders. Most offenses involving deepfakes haven't been criminalized yet. As noted above, legal options for victims of nonconsensual deepfake pornography are limited.

The newness of deepfakes also presents a challenge to the court system. Explaining deepfake-related crimes to judges is difficult. An expert providing training to practitioners, such as judges, mental health professionals, law enforcement officials, and educators on this issue noted that 80% of the training participants "have no idea what a deepfake is."¹⁵ There is thus little fear of prosecution.³¹

Recent advances in AI have made it easy to engage in intrinsically as well as financially motivated offenses using deepfakes. Most victims have no guardrails against fraud involving deepfakes. Guardrails that consumers use to protect against cybercriminals do not provide an adequate means of protection for deepfakes. Integrative approaches combining policy and technological measures at various levels are needed to fight deepfakes. 

REFERENCES

1. "Deepfakes: How it all began – And where it could lead us." The Decoder. Accessed: Apr. 28, 2022. [Online]. Available: <https://the-decoder.com/history-of-deepfakes/>
2. D. Byman et al., "The deepfake dangers ahead," *Wall Street J.*, Feb. 2023. Accessed: May 5, 2023. [Online]. Available: <https://www.wsj.com/articles/the-deepfake-dangers-ahead-b08e4ecf>
3. "Increasing threat of deepfake identities," U.S. Department of Homeland Security, Washington, DC, USA, Jul. 6, 2019. [Online]. Available: <https://www.dhs.gov/sites/default/files/>

publications/increasing_threats_of_deepfake_identities_0.pdf

4. V. Kropotov et al. "How underground groups use stolen identities and deepfakes." Trend Micro. Accessed: Sep. 27, 2022. [Online]. Available: https://www.trendmicro.com/en_us/research/22/i/how-underground-groups-use-stolen-identities-and-deepfakes.html
5. S. Dunn, "Women, not politicians, are targeted most often by deepfake videos," Centre for International Governance Innovation, Waterloo, ON, Canada, Mar. 2021. [Online]. Available: <https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/>
6. C. Klein, "'This will be dangerous in elections': Political media's next big challenge is navigating Ai deepfakes," *Vanity Fair*, Mar. 2023. Accessed: May 5, 2023. [Online]. Available: <https://www.vanityfair.com/news/2023/03/ai-2024-deepfake>
7. J. J. Low. "Are any of us safe from deepfakes?" TechHQ. Accessed: Aug. 28, 2020. [Online]. Available: <https://techhq.com/2020/08/are-any-of-us-safe-from-deepfakes/>
8. J. Bateman, "Get ready for deepfakes to be used in financial scams," Carnegie Endowment for International Peace, Washington, DC, USA, Aug. 2020. [Online]. Available: <https://carnegieendowment.org/2020/08/10/get-ready-for-deepfakes-to-be-used-in-financial-scams-pub-82469>
9. A. Romano. "Deepfakes are a real political threat." Vox. Accessed: Oct. 7, 2019. [Online]. Available: <https://www.vox.com/2019/10/7/20902215/deepfakes-usage-youtube-2019-deeptrace-research-report>
10. R. DePompa and D. Molina. *Swapped Out: Hackers Target Social Media Users with High-Tech Fake Videos*. (May. 2022). [Online Video]. Available: <https://www.wsaz.com/2022/05/16/swapped-out-hackers-target-social-media-users-with-high-tech-fake-videos/>

11. B. Frey, *Not Just for the Money: An Economic Theory of Personal Motivation*. Northampton, MA, USA: Edward Elgar Publishing, 1997.
12. E. L. Deci and R. M. Ryan, *Intrinsic Motivation and Self-Determination*. New York, NY, USA: Human Behavior Plenum Press, 1985.
13. R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *Amer. Psychologist*, vol. 55, no. 1, pp. 68–78, Jan. 2000, doi: 10.1037/0003-066X.55.1.68.
14. G. Patrini, H. Ajder, and F. Cavalli. "Automating image abuse: Deepfake bots on telegram." GitHub. Accessed: Oct. 20, 2020. [Online]. Available: <https://giorgiop.github.io/posts/2020/10/20/automating-image-abuse/>
15. K. Hao. "Deepfake porn is ruining women's lives." MIT Technology Review. Accessed: Feb. 12, 2021. [Online]. Available: <https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/>
16. M. Burgess. "A deepfake porn bot is being used to abuse thousands of women." Wired. Accessed: Oct. 20, 2020. [Online]. Available: <https://www.wired.co.uk/article/telegram-deepfakes-deepnude-ai>
17. E. Borgos. "How I made a deepfake of Elon Musk: One man's journey into the world of AI-generated videos." Medium. Accessed: May 17, 2020. [Online]. Available: <https://medium.com/swlh/how-i-made-a-deepfake-of-elon-musk-7b1aae06fe01>
18. Z. Shapira, "Expectancy determinants of intrinsically motivated behavior," *J. Pers. Social Psychol.*, vol. 34, no. 6, pp. 1235–1244, Dec. 1976, doi: 10.1037/0022-3514.34.6.1235.
19. J. Kahn, "Here's who created those viral Tom Cruise deepfake videos," *Fortune*, Mar. 2021. Accessed: May 5, 2023. [Online]. Available: <https://fortune.com/2021/03/02/tom-cruise-deepfake-videos-tik-tok-chris-ume/>
20. J. Rohrlisch, "Romance scammer used deepfakes to impersonate a navy admiral and bilk widow out of nearly \$300,000," *The Daily Beast*, Oct. 2020. Accessed: May 5, 2023. [Online]. Available: <https://www.thedailybeast.com/romance-scammer-used-deepfakes-to-impersonate-a-navy-admiral-and-bilk-widow-out-of-nearly-dollar300000>
21. N. Kshetri, *Cybersecurity Management: An Organizational and Strategic Approach*. Toronto, OH, USA: The Univ. of Toronto Press, 2021.
22. L. Mcguire. "Manga artist scammed half a million dollars to Mark Ruffalo deepfake." Screen Rant. Accessed: Sep. 24, 2022. [Online]. Available: <https://screenrant.com/mark-ruffalo-scam-artist-manga-deepfake-marvel/>
23. M. Takahashi. "Manga artist falls for fake Mark Ruffalo, loses \$500,000." *The Asahi Shimbun*. Accessed: Sep. 22, 2022. [Online]. Available: <https://www.asahi.com/ajw/articles/14722566>
24. M. Anderson. "Deepfaked voice enabled \$35 million bank heist in 2020mm." *Unite.AI*. Accessed: Oct. 15, 2021. [Online]. Available: <https://www.unite.ai/deepfaked-voice-enabled-35-million-bank-heist-in-2020/>
25. "China's deepfake anchors spread disinformation on social media," *Radio Free Asia*, Washington, DC, USA, Feb. 2023. [Online]. Available: <https://www.rfa.org/english/news/china/china-deepfake-02082023032941.html>
26. S. Thompson, "Making deepfakes gets cheaper and easier thanks to A.I.," *NYTimes*, Mar. 2023. Accessed: May 5, 2023. [Online]. Available: <https://www.nytimes.com/2023/03/12/technology/deepfakes-cheapfakes-videos-ai.html>
27. "Cyber goons extorting with deepfake porn clips," *The Times of India*, Mar. 2021. Accessed: May 5, 2023. [Online]. Available: <https://timesofindia.indiatimes.com/city/allahabad/cyber-goons-extorting-with-deepfake-porn-clips/articleshow/81537505.cms>
28. M. Westerlund, "The emergence of deepfake technology: A review," *TIM Review*, Ottawa, ON, Canada, Nov. 2019. Accessed: May 5, 2023. [Online]. Available: <https://timreview.ca/article/1282>
29. A. Tadepalli, "In the age of information, can we weed out the fake news?" Cal Alumni Association, Berkeley, CA, USA, Aug. 2020. [Online]. Available: <https://alumni.berkeley.edu/california-magazine/online/age-information-can-we-weed-out-fake-news/#>
30. B. Fowler. "Deepfakes pose a growing danger, new research says." *CNET*. Accessed: Aug. 8, 2022. [Online]. Available: <https://www.cnet.com/tech/services-and-software/deepfakes-pose-a-growing-danger-new-research-says/>
31. A. Court, "Twitch star QTCinderella's deepfake porn nightmare: 'F-k the internet'," *NY Post*, Feb. 2023. Accessed: May 5, 2023. [Online]. Available: <https://nypost.com/2023/02/06/twitch-star-tearfully-reveals-shes-victim-of-deepfake-porn-f-k-the-internet/>
32. "India's scam central: Inside villages of cyber cheats," *The Statesman*, Jul. 2022. Accessed: May 5, 2023. [Online]. Available: <https://www.thestatesman.com/india/indias-scam-central-inside-villages-cyber-cheats-1503093010.html>
33. V. Karasavva and A. Noorbhai, "The real threat of deepfake pornography: A review of Canadian policy," *Cyberpsychol., Behav., Social Netw.*, vol. 24, no. 3, pp. 203–209, Mar. 2021, doi: 10.1089/cyber.2020.0272.

NIR KSHETRI is a professor at the University of North Carolina at Greensboro, Greensboro, NC 27412, USA, and the "Computing's Economics" column editor of *Computer*. Contact him at nbkshetr@uncg.edu.