QRBD

Quarterly Review of Business Disciplines

February 2025

Volume 11
Number 3/4

A Journal of International Academy of Business Disciplines
Sponsored by University of North Florida
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)

QRBD - QUARTERLY REVIEW OF BUSINESS DISCIPLINES

A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

The *Quarterly Review of Business Disciplines (QRBD)* is published by the International Academy of Business Disciplines in February, May, August, and November.

Manuscript Guidelines/Comments. *QRBD* is a blind peer-reviewed journal that provides publication of articles in all areas of business and the social sciences that affect business. The Journal welcomes the submission of manuscripts that meet the general criteria of significance and business excellence. Manuscripts should address real-world phenomena that highlight research that is interesting and different – innovative papers that begin or continue a line of inquiry that integrate across disciplines, as well as, those that are disciplinary. The Journal is interested in papers that are constructive in nature and suggest how established theories or understandings of issues in business can be positively revised, adapted, or extended through new perspectives and insights. Manuscripts that do not make a theoretical contribution to business studies or that have no relevance to the domain of business should not be sent to *QRBD.* Submissions to *QRBD* must follow the journal's Style Guide for Authors, including length, formatting, and references. Poorly structured or written papers will be returned to the authors promptly. Manuscript length is approximately 16 – 20 single-spaced pages. Acceptance rate is 30%.

Description. The *Quarterly Review of Business Disciplines* is a quarterly publication of the International Academy of Business Disciplines (IABD); a worldwide, non-profit organization established to foster and promote education in all of the functional and support disciplines of business. The objectives of *QRBD* and IABD are to stimulate learning and understanding and to exchange information, ideas, and research studies from around the world. The Academy provides a unique global forum for professionals and faculty in business, communications, and other social science fields to discuss and publish papers of common interest that overlap career, political, and national boundaries. *QRBD* and IABD create an environment to advance learning, teaching, and research, and the practice of all functional areas of business. *Quarterly Review of Business Disciplines* is published to promote cutting edge research in all of the functional areas of business.

Submission Procedure. **Manuscripts should be uploaded to Easy Chair using the link provided on our website**. Manuscripts must follow the submission guidelines, **including all formatting requests**, and be well edited. Upon completion of a review by expert scholars who serve on the *QRBD* Editorial Review Board, the first author will be informed of acceptance or rejection of the paper within a one to two-month timeframe from the submission date. If the paper is accepted, the first author will receive a formal letter of acceptance along with the *QRBD* Style Guide for Authors.

IABD members and authors who participate in the IABD annual conference are given first priority as a matter of courtesy. For additional information, please visit www.iabd.org.

QRBD - QUARTERLY REVIEW OF BUSINESS DISCIPLINES

A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

MISSION STATEMENT

The organization designated as the International Academy of Business Disciplines is a worldwide, non-profit organization, established to foster and promote education in all of the functional and support disciplines of business.

WWW.IABD.ORG
WWW.QRBD.NET

The Quarterly Review of Business Disciplines (QRBD) is listed in *Cabell's Directory of Publishing Opportunities.*

QRBD - QUARTERLY REVIEW OF BUSINESS DISCIPLINES

A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

VOLUME 11, NUMBER 3/4

February 2025

ISSN 2329-5163 (online)

CONTENTS

Your De-Boarding Group Has Been Called:

Maintaining Dignity within Employee Terminations

Vance Johnson Lewis, Northeastern State University

As this eleventh volume of *QRBD* comes to a conclusion, the world in which we research and publish is rapidly changing. While in this issue we explore issues related to artificial intelligence and deep-fakes, why people do or do not respond to surveys, and the history of one of the giants of the entertainment industry, around us we continue to see the dissolution of academic institutions, the destruction of diversity, equity, and inclusion programs, the restructuring of both the United States government and our relationships with our allies, and the continued closures of once prominent retail staples. We have even seen the historic first round of layoffs in the traditionally people centric Southwest Airlines.

While my eyes have been focused on many of these mentioned changes, it is the latter that particularly struck home to me. While Director of Organizational Behavior and Human Resources at The University of Texas at Dallas, my students and I enjoyed a strong relationship with Southwest Airlines, with many of my students finding internships and permanent employment with this organization. Known for being the "airline with a heart", on February 17, 2025, this organization which subscribed to the philosophy that happy employees make happy customers abruptly sent their employees home from their Dallas, TX, based headquarters with the knowledge that the next morning, 15% of its corporate workforce would be terminated (Singh, 2025). Bowing to apparent pressure from investors, this company who had never enacted a major layoff in its 54 years history (Snider, 2025), the airline with a heart suddenly appeared to be heartless.

Terminating with dignity

As my alumni face their employment loss, I reflected on my own experiences with termination. While thankfully it has only occurred twice in my life, I think on how I was (not) treated with dignity… the idea that all people have a basic worth and status that gives rise to fundamental rights and respect. The management of employee terminations represents a significant challenge for organizations, respecting individual dignity during this often painful process is not simply a question of organizational policy; it serves as an artifact of the company culture and the value that employers place on employees within the workplace (Lucas et al., 2017). The way in which the organizations manage terminations reflects on their corporate ethos and can have an impact not only on the people who are directly impacted, but also on the morales of the remaining employees. When dignity is supported, both remaining and terminated employees are likely to maintain a sense of self-esteem, even during difficult transitions, thus promoting a constructive organizational culture.

Dignity in the workplace includes the recognition and respect for the intrinsic value of each employee, which is significant in maintaining a positive organizational environment. Lucas (2017) stresses that dignity in the workplace is essential to mitigate the psychological impacts of layoffs. Employees who feel appreciated and respected are more likely to have better emotional

well-being, higher levels of commitment and a greater sense of loyalty towards their organization, even if they must leave due to wider organizational changes. Inversely, negative termination experiences can lead to long-lasting resentments, disengagement on the part of the remaining staff and a clouded brand image, which can have vast implications for the company's ability to both recover from and move forward from the termination.

Facing the change

The implications of organizational change are essential to understanding the context that surrounds terminations. As organizations evolve, employees often find themselves navigating in turbulent waters. D'Cruz et al. (2014) note that the challenges posed by organizational change can exacerbate feelings of insecurity, anxiety, and uncertainty among employees. These emotions can be particularly powerful during layoffs, since individuals face the potential loss of income and their professional identities. Addressing these emotional and psychological challenges in a significant and ethical way is essential for organizations that aim to maintain dignity during the termination process.

A critical component of facing the changes brought by terminations is effective communication. The ways in which organizations transmit the termination news can be fundamental in how those affected process the change. According to Lucas (2015) and Noronha et al. (2020), transparent communication strategies help mitigate negative psychological impacts associated with loss of employment. When organizations proactively communicate the reasons for the termination, they describe the available support and express empathy towards the situation of the affected employee, thus promoting the terminated employee's sense of worth and minimizing feelings of abandonment and isolation. This is particularly important as the emotional consequences of terminations are felt not only by the terminated but also by the remaining employees. By promoting a communicative environment where honest discourse thrives, organizations can maintain a level of dignity and respect for the affected workforce along with minimizing any type of resentment or survivor's guilt felt by those not affected.

The timeliness of the termination is also crucial to honoring the dignity of the affected. Rumors of layoffs have been found to have a profound impact on the stress levels of employees (Cohen, 1995). Of course, workplace hindrance stressors have been shown to negatively impact job performance as well as organizational commitment, leading to increased anxiety among the remaining employees, creating a toxic environment, where employees feel insecure and undervalued (Baker & Lucas, 2017). While laws, such as Worker Adjustment and Retraining Notification (WARN) Act, dictate how much advance notice an employee should receive prior to termination, immediate notification that a layoff is going to occur is crucial to maintaining not only the dignity of the employee but also their mental well being. One of the worst experiences an employee can face is the "mystery meeting" when an unexpected meeting with management or human resources is scheduled with no explanation, causing days of anxiety and stress. Managers should avoid this delay in favor of scheduling meetings as expediently as possible with the clearly communicated message that the meeting is related to the continuation of employment as to embrace the well being of the employee as well as avoid any ripple affects the termination might have.

Avoid the box

Perhaps nothing symbolized a more ineffectively managed termination than the employee being handed a box (made worse by a security guard/police officer) and being marched out of the building. To effectively implement ethical termination practices, organizations must develop a structured model that includes comprehensive training for managing the principles of dignity and respect in employment. This training should incorporate an understanding of human resource laws and practices within the workplace, ensuring that all staff are equipped to deal with the terminations with the sensitivity and equity they require (McDougal et al., 2018; Grandy & Mavin, 2017). A suggested model can include training on active listening, empathic decision making and directive communication, allowing managers to get involved with employees about their terminations in an attentive and constructive way.

Aside from simply delivering the news, termination models should include career counseling services for affected employees, which demonstrates a tangible commitment to their future employment. Beyond simple compensation practices, special attention should be given to resources that help employees find new positions and process their experiences in a healthy way. By promoting a culture that values each individual's contributions and recognizes the complexities that involve employment transitions, organizations can cultivate an environment in which dignity prevails, positioning itself as leaders in ethical employment practices. The premise of promoting support routes for affected employees is based on the belief that dignity and respect must even be maintained, even after termination has occurred (Wieland, 2020).

Taking Flight after Termination

On February 18, 2025, Southwest Airlines proceeded with their mass layoffs. In an email sent to the Texas Workforce Commission, Southwest's Vice President of People Lindsey Lang offered, in keeping with the WARN Act, a list of 626 positions that would be eliminated from the Dallas headquarters, which did include the Director of Diversity, Equity, and Inclusion (Lang, 2025). The email assured that those affected would receive pay and benefits through April 22, 2025, provided they agree to the offered separation agreement. While Southwest did employ fairly good practice in notifying these affected individuals in an expedient manner and offering some compensation packages, the media blitz along with the still unidentified (as of this editorial) 1000 positions leaves Southwest open to potential pitfalls of a mishandled termination situation. While not conducted completely out of sync with their people-first values, only time will tell if Southwest employees and customers will see if they were treated with dignity or if Southwest has irreparably damaged their reputation through mishandled terminations.

References

Baker, S. J., & Lucas, K. (2017). Is it safe to bring myself to work? Understanding LGBTQ experiences of workplace dignity. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration, 34*(2), 133-148.

Cohen, L. (1995). Looming public-service layoffs causing stress, other problems, Ottawa MDs report. *CMAJ: Canadian Medical Association Journal*, *152*(11), 1891.

D'Cruz, P., Noronha, E., & Beale, D. (2014). The workplace bullying-organizational change interface: Emerging challenges for human resource management. *The International Journal of Human Resource Management, 25*(10), 1434-1459.

Lang, Lindsey. Email/Letter offered to Texas Workforce Commission, 18 Feb. 2025.

Lucas, K., Manikas, A. S., Mattingly, E. S., & Crider, C. J. (2017). Engaging and misbehaving: How dignity affects employee work behaviors. *Organization Studies, 38*(11), 1505-1527.

Lucas, K. (2017). Workplace dignity. *The international encyclopedia of organizational communication, 4*, 2549-2562.

Lucas, K. (2015). Workplace dignity: Communicating inherent, earned, and remediated dignity. *Journal of Management Studies, 52*(5), 621-646.

McDougal, M. S., Lasswell, H. D., & Chen, L. C. (2018). *Human rights and world public order: the basic policies of an international law of human dignity*. Oxford University Press.

Noronha, E., Chakraborty, S., & D'Cruz, P. (2020). 'Doing dignity work': Indian security guards' interface with precariousness. *Journal of Business Ethics, 162*(3), 553-575.

Singh, R. K. (2025, Feb. 20). Southwest's layoffs dent its worker-first culture. *Reuters*. Retrieved from https://www.reuters.com/business/aerospace-defense/southwests-layoffs-dent-its-worker-first-culture-2025-02-19/

Snider, R. (2025, Feb. 20). Who's getting laid off at Southwest Airlines? Here are the jobs we know. *WFAA.* Retrieved from https://www.wfaa.com/article/money/business/southwest-airlines-layoffs-jobs-how-many-which-positions-dallas-texas/

Wieland, S. M. (2020). Constituting resilience at work: Maintaining dialectics and cultivating dignity throughout a worksite closure. *Management Communication Quarterly, 34*(4), 463-494.

An analysis of item non-response in a survey of law students, attorneys, and judges

Catherine Lau Crisp, University of Arkansas at Little Rock

Kaye McKinzie, University of Central Arkansas

Candace McCown, Arkansas Judges and Lawyers Assistance Program

Jennifer Donaldson, Arkansas Judges and Lawyers Assistance Program

Abstract

Missing data is an important issue in summed scales used in survey research. It can have a significant impact on the quality of the research by reducing the usable sample size, reducing the statistical power in small samples, limiting the generalizability of the results, and forcing the researcher to make decisions about whether to exclude responses from the analysis or to use a data replacement method. Excluding responses from the analysis reduces the usable size while replacing the missing items may result in an overestimation or underestimation of scale scores, affect the measure's reliability, and increase the likelihood of finding statistically significant results when there are none. Despite these challenges, missing data is rarely the focus of research studies. This article focuses on missing data in a sample of lawyers and law students who completed a survey that consisted of six different summed scales, each of which required that respondents answer all questions in the measure to compute an accurate score. The questions of interest in this study were 1) whether any demographic groups or combinations of groups were more or less likely to respond to all items in the summed scales and 2) whether there were statistically significant relationships between respondents' willingness to complete all items in the different summed scales. Implications for further research are discussed.

Keywords: missing data, item non-response, law students, summed scales, surveys, judges, attorneys

An analysis of item non-response in a survey of law students and attorneys

Introduction

The use of summed scales in research has a long history. Its onset is frequently attributed to Rensis Likert, who developed the widely used and accepted Likert scale in 1932 and is credited with its origins. Spector (1992) states that summed scales have four characteristics:

1. The scale must contain multiple items that will be combined or summed.
2. Each item must measure something that varies quantitatively.
3. Each item has no "right" answer.
4. Each item must be phrased as a statement in which respondents are asked to give a rating that best reflects their response to the item.

One issue that arises in research using summed scales is the issue of missing values as all questions in the scale must be answered in order to compute a score that reflects the respondents' true feelings about the issue of interest. When respondents neglect to respond to items in a summed scale (referred to as item non-response), the researcher has two choices: either delete the response using one of two methods (listwise or pairwise) or replace the data using one of several methods. Decisions about whether to delete or replace the data are often made based on a variety of factors including the extent of the missing data and whether the data can be characterized as Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). Deleting individual cases in which a single data point is missing (listwise deletion) can significantly reduce to pool of usable data while replacing the data can artificially increase scores and may not reflect respondents' actual feelings about the construct of interest.

Given the aforementioned issues, the questions of interest in this analysis were:

1. Are any demographic groups or combinations of groups more or less likely to respond to all items in the summed scales and
2. Are there statistically significant relationships between respondents' willingness to complete all items in the different summed scales?

Causes and implications of missing data

Item non-response is the failure of the respondent to answer individual items in a survey, despite being eligible to respond. According to de Leeuw, Hox, and Hussman (2003), item non-response has three different forms:

1. Information that is not provided by the respondent for certain question(s);
2. Information that is provided but is not usable to the researchers; and
3. Information that was provided but is lost and cannot be retrieved

Moreover, item non-response may occur when the items are missing by design, items do not apply to the respondent, the respondent has difficulty with the cognitive task involved in answering the question, the respondent refuses to respond, the respondent does not know the answer, and the respondent provides a response that does not fit the given response categories (Hussman, 1999). Item non-response may also be influenced by the sensitive nature of questions

such as questions about income, sexual behaviors, and drug and alcohol use (Blair, Sudman, Bradburn, & Stocking, 1977; Chen, 2019).

Item non-response can have a significant impact on the quality of the research. It may reduce the usable sample size, reduce the statistical power in small samples, limit the generalizability of the results, and force the researcher to make decisions about whether to exclude responses from the analysis or to use a data replacement method (McNeeley, 2012). Excluding responses from the analysis reduces the usable size while replacing the missing items may result in an overestimation or underestimation of scale scores, affect the reliability of the measure, and increase the likelihood of finding statistically significant results when there are none.

Methodology

Sample

This study sampled approximately 800 law students and 10,000 attorneys and judges licensed in a southern state in the United States. Its purpose was to examine mental health and well-being among this population and compare the findings to other studies conducted on this population.

IRB approval and survey distribution

Before beginning this study, approval was obtained from the first author's institutional review board (IRB). Once the approval was granted, the research term began data collection. Prior to beginning this research, the research team had concerns about asking law students, attorneys, and judges about sensitive issues such as their mental health and substance use. McNeeley (2012) claims that while there is no clear definition of sensitive issues, it is generally agreed that drug or alcohol use is considered a sensitive topic. Tourangeau and Smith (1996) state that a survey question is "sensitive if it raises concerns about disapproval or other consequences (such as legal sanctions) for reporting truthfully or if the question itself is seen as an invasion of privacy'' (p. 276). McNeeley (2012) further states that the characteristics of the interest group must be considered as the sensitivity of the topic can vary by the target population. Because the questions inquired about licensed law students', lawyers', and judges' use of drugs and alcohol and their experience of depression, anxiety, and stress, several steps were taken to protect respondents' confidentiality. First, the recruiting email for the survey was sent from a member of the research team to a member of the licensing board who then sent the recruiting email to the approximately 10,000 licensed attorneys and judges. A separate email was sent to contacts in each of the law schools who then sent the survey to students enrolled in their respective programs. No member of the research team had access to any contact information for the law students, attorneys, and judges to whom the survey was sent. Second, all data was stored on an encrypted server owned by the Arkansas Judges and Lawyers Assistance Program. As such, the data is legally protected from being subpoenaed or requested by outside sources (Rule 10, 2017). Third, the above information was made explicit in the informed consent that respondents agreed to prior to beginning the survey. Fourth, because the IRB was concerned that requiring respondents to respond to any questions after agreeing to the informed consent could be seen as coercive, respondents simply had to agree to the informed consent and could submit the survey without answering any questions in the survey in order to receive the ability to earn

one Continuing Legal Education (CLE) credit at no cost. Finally, categories in the demographic variables with less than 10 respondents were combined into a single category and the research team only reported categories with 10 or more respondents. These measures were intended to obtain more truthful answers to the sensitive questions asked in this survey.

Following approval by the IRB, the state board that oversees licensed attorneys, and the deans of the law schools, emails were sent on April 9, 2024, inviting law students, attorneys, and judges to participate in the study. Reminder emails were sent via the licensing board and law schools every two weeks until June 18, 2024. In addition, information about the survey was distributed via social media and at conferences and judicial meetings during the same dates.

Instruments

Given that the focus of this study was on the well-being of law students and licensed attorneys and judges in a southern state in the United States, demographic questions and four summed scales that examined mental health and well-being were included in the survey packet. Three of the four measures (AUDIT, DASS-21, and DAST-10) and the demographic questions were asked to facilitate comparison to larger studies that had been conducted on well-being among attorneys and judges by Krill and colleagues (Anker & Krill, 2021; Krill, Johnson, & Albert, 2016). The fourth measure (MSC-SF) was included to examine the use of self-compassion in attorneys and judges and to expand the body of research on well-being among this population. Each of the instruments used in this study had been used in several other studies, had strong internal consistency, and undergone extensive validation studies.

Depression Anxiety Stress Scales (DASS-21)

The 21-item DASS-21 (Henry & Crawford, 2005) is a shorter version of the 42-item DASS-42 developed by Loviband and Loviband (1995) to assess depression, anxiety, and stress, each of which is measured with a separate subscale and is analyzed as a separate scale. Scores for each subscale range from 0 - 21, with higher scores indicating more severe symptoms.

Alcohol Use Disorders Identification Test (AUDIT)

The 10-item instrument was developed by the World Health Organization to screen for unhealthy alcohol use (Saunders, Aasland, Babor, De La Fuente, & Grant, 1993). Scores range from 0 - 40 with higher scores reflecting more hazardous alcohol consumption.

Drug Abuse Screening Test (DAST-10)

The 10-item DAST-10 was developed to assess drug use, not including alcohol and tobacco use, in the past 12 months. Scores range from 0 - 10 with higher scores reflecting a severe level of use (Skinner, 1982).

Mindful Self-Compassion Scale-Short Form

The 12-item Mindful Self-Compassion Scale-Short Form (MSC-SF) (Raes, Pommier, Neff, & Van Gucht, 2011) is a shorter version of the 26-item Mindful Self-Compassion Scale developed by Neff (2003) to assess self-compassion. Scores on the MSC-SF range from 1 - 5 with higher scores reflecting more self-compassion.

Statistical Analysis

The analysis was conducted using SAS Studio and Microsoft Excel 365. Responses to the demographic questions, each of the three subscales of the DASS-21 were examined individually along with the AUDIT, DAST-10, and SCS-SF. As noted earlier, to protect the confidentiality of the respondents as much as possible, demographic characteristics with less than 10 respondents were combined with other categories in the same variable so that only categories with 10 or more responses were reported. In examining the scale items, if a single item on one of the summed scales was missing, the score could not be calculated and the scale was thus coded as 0. When all the items in the scale were answered, a total score could be calculated and the scale was coded as 1. Chi-square, correlation tests, missing in common, and logistic regression were analyzed to assess which groups were more or less likely to answer all items on each of the six measures of interest and to examine relationships between the demographic variables and completion of the summed scales.

Results

Responses to summed scales

Of the 1,547 surveys returned, 37 (2.4%) were returned with only the informed consent completed. This resulted in 1,510 (97.6%) surveys in which the respondents answered at least one question beyond the informed consent and are the focus of our analysis. 1,149 (76%) respondents answered all 53 items in the summed scales while 1,425 answered all the demographic questions (94.37%). This resulted in 1,121 (74%) who answered all seven of the demographic measures.

Table 1

Missing data in summed scales

	Answered		Missing	
N = 1,510	*n*	%	*n*	%
DASS-21 (full scale)	1,439	95.3	71	4.7
Depression	1,483	98.2	27	1.8
Anxiety	1,490	98.7	20	1.3
Stress	1,479	98	31	2.1
AUDIT	1,288	85.3	222	15
DAST-10	1,427	94.5	83	5.5
SCS-SF	1,424	94.3	86	5.7

Demographics

Personal Characteristics

There was very little missing data in respondents' personal characteristics with a range of 0.27% (gender) to 1.06% (whether respondents had children). The sample consisted of more women than men and more Caucasian/white than other races. The most common age group was 41 - 50. Approximately two-thirds of the respondents had children and almost the same percent were married.

Table 2

Personal demographic characteristics

Characteristic (N = 1,510)	*n*	%
Gender		
Female	757	50.13
Male	737	48.8
Other	12	0.8
Did not answer	4	0.27
Race		
Black or African American	55	3.64
Caucasian/white	1,347	89.2
Multiracial	69	4.57
Single race other than African American or Caucasian	25	1.66
Did not answer	14	0.93
Age		
30 or younger	196	12.98
31 - 40	331	21.92
41 - 50	353	23.37
51 - 60	348	23.05
61 - 70	200	13.25
71 or older	68	4.5
Did not answer	14	0.93
Did you have children?		
Yes	997	66.03
No	497	32.91
Did not answer	16	1.06
Marital status		
Married	1011	66.96

Partnered but not married	79	5.23
Divorced	162	10.73
Separated	13	0.86
Single	213	14.1
Widowed	25	1.66
Did not answer	7	0.46

Notes:
1. This category includes respondents who identified as non-binary and others who identified in other ways. Because the N for each of these groups is less than 10, they are subsumed into a single category.
2. This category includes all respondents who checked more than one racial category.
3. This category includes all respondents who identified with a single race other than White/Caucasian or Black/African-American. Because the N for each of these groups is less than 10, they are subsumed into a single category.

Professional characteristics

There was less missing data for the professional characteristics than for the personal characteristics with just 1 (0.06%) respondent neglecting to answer their role and 3 (0.2%) respondents neglecting to answer the number of years they had worked in law. Almost 25% of respondents had worked in the legal field for 11 - 20 years while Managing Partner was the most frequently cited role with 26.76% of respondents identifying this as their role.

Table 3

Professional demographic characteristics

Characteristic (N = 1,510)	*n*	%
Years in field		
0 - 10	593	39.27
11 - 20	369	24.43
21 - 30	303	20.07
31 - 40	180	11.92
41 or more	62	4.11
Did not answer	3	0.2
Role		
Judge	55	3.64
Senior Partner	215	14.23
Senior Associate	97	6.42
Managing Partner	404	26.76
Junior Partner	43	2.85
Junior Associate	102	6.76

Law school faculty/staff	15	0.99
Law school student	76	5.04
Staff attorney/clerk/paralegal	215	14.24
None of the above	287	19.01
Did not answer	1	0.06

Likelihood of not responding to all items by demographic characteristic

Gender

There were significant differences in responses to five of the six measures as women and men were significantly more likely than "other" to complete all items in the three subscales of the DASS-21, AUDIT, and DAST-10 (Chi Sq: DASS-21 depression = 0.0042; DASS-21 anxiety = 0.001; DASS-21 stress = 0.0070; AUDIT = 0.0032; DAST-10 < 0.001). A small but insignificant difference with gender was found in the respondents' completion of the SCS-SF.

Race

There were significant differences in responses to three of the six measures (Chi Sq: DASS-21 depression <0.001; DASS-21 anxiety = 0.0108; DAST =0.0078). Those who did not answer the question were significantly less likely than other racial groups to complete all items on the DASS-21 depression subscale and DAST-10; they were significantly less likely than groups other than those with single race other that was not Caucasian/White or African-American/Black to complete all items on the DASS-21 anxiety subscale. Those identified with a single race that was not Caucasian/White or African-American/Black were significantly less likely than all other groups to answer all items on the DASS-21 anxiety subscale.

Age

There were significant differences in the completion of the scales by age. Those who identified as 71 or older were significantly less likely to answer all items in five of the six scales (Chi Sq: DASS-21 depression = 0.0034; DASS-21 anxiety = 0.0326; AUDIT = 0.0143; DAST < 0.001; SCS-SF < 0.001) but were not significantly less likely to complete the DASS-21 stress subscale.

Marital status

Respondents who did not answer this question were the least likely to answer the DASS-21 anxiety subscale followed by those who were separated (Chi Sq: DASS-21 anxiety = 0.0128).

Have children

Respondents who did not answer if they had children were less likely to answer the DASS-21 Depression measure (Chi Sq: DASS-21 depression = 0.005).

Years as an attorney

Respondents who did not answer the number of years they have worked as an attorney and those who have worked for 41 years or more as an attorney were less likely to complete five of the six measures (Chi Sq: DASS-21 depression < 0.001; Anxiety < 0.001; Stress = 0.0021; DAST < 0.001; SCS < 0.001;).

Role

There were significant differences in respondents' roles in the completion of four of six measures. Law school faculty/staff were the least likely to answer all items in the DASS-21 depression and SCS-SF. Judges were the least likely to answer all items in the AUDIT and DAST-10 (Chi Sq: DASS-21 depression = 0.0182; AUDIT = 0.0007; DAST-10 = 0.015; SCS-SF = 0.0003).

Missing data in common

An analysis of the patterns of missing data was conducted to assess patterns of missingness. This is often referred to as a missing in common qualitative analysis. We looked for three different patterns of missing in common data:

1. Relationships between demographic questions: When respondents failed to answer a demographic question, were there patterns in their failure to answer other demographic questions?
2. Relationships between summed scales: When respondents failed to answer a single item on one summed scale, were there patterns in their failure to answer items on other scales in the survey?
3. Relationships between demographic questions and summed scales: When respondents failed to answer items on demographic questions, were there patterns in their failure to answer items on the summed scales? Correspondingly, when respondents failed to answer items on summed scales, were there patterns in their failure to answer demographic questions?

Demographic variables

An examination of missing in common demographic characteristics (failure to answer demographic questions) found 10 different combinations in which respondents failed to answer more than one demographic question and usually consisted of one respondent per combination. However, the largest instance was in only three cases (respondents failed to answer questions regarding race, age, marital status, and whether they had children). This small sample size precluded further analysis and/or conclusions about the pattern.

Summed scales

As previously noted, the failure to answer a single item on a summed scale renders the response to that scale unusable or leads the researcher to use a data replacement method. Given

this, we assessed whether there were patterns in respondents' willingness to respond to all items in different combinations of the summed scales. We found that:

1. 45 respondents failed to complete at least one item in both the AUDIT and the DAST-10;
2. 17 respondents failed to answer at least one item on both the AUDIT and SCS-SF.
3. All other combinations of scales in which respondents failed to complete at least one item in multiple scales had 7 respondents or less (small sample size).

Following the above analysis, further analysis was conducted to assess whether there was a relationship between respondents' completion of all items on the different summed scales. Weak but significant correlations exist between respondents' completion of the three subscales (depression, anxiety, and stress) of the DASS-21: respondents who completed all items in one measure were more likely to complete all items in the other two measures. This is to be expected since they are subscales of a single scale and are presented as a single measure in the survey instrument.

Weak but significant correlations were also found in respondents' completion of the AUDIT and SCS-SF and the DAST-10 and SCS-SF. The strongest correlation found was a moderate but significant relationship between the AUDIT and DAST-10. This is an interesting finding since the AUDIT asks about alcohol use and the DAST-10 asks about drug use, both of which are considered sensitive topics. This finding may indicate that those who are willing to answer questions about alcohol use are also willing to answer questions about drug use and that the converse is also true.

It is interesting to note and makes sense that the strongest relationship (AUDIT & DAST-10) is the one that had the most respondents (45) with this missing in common pair.

Table 4

Correlation of scale completion (N = 1,510)

Variable	DASS-21 depression	DASS-21 anxiety	DASS-21 stress	AUDIT	DAST-10	SCS-SF
	α	*α*	*α*	*α*	*α*	*α*
DASS-21 depression	-	**0.1155***	**0.1214***	-0.0136	0.0113	0.0746
DASS-21 anxiety	**0.1155***	-	0.024	0.0173	0.0228	-0.0034
DASS-21 stress	**0.1214***	0.024	-	0.0337	-0.0144	0.045
AUDIT	-0.0136	0.0173	0.0337	-	**0.3429***	**0.1077***
DAST-10	0.1131	0.0228	-0.0144	**0.3429***	-	**0.1162***
SCS-SF	0.0746	-0.0034	0.045	**0.1077***	**0.1162 ***	-

*$p < .001$.

Demographic variables and summed scales

An analysis of the relationship between missing values in demographic variables and summed scales found no significant combinations. There were only two combinations where more than one respondent had the same combinations of missing demographic variables and incomplete summed scales (age and the DAST-10 ($N = 3$); having children and the AUDIT ($N = 2$)). These small samples prevented further analysis.

Regression model

Logistic regression was used to assess whether any of the seven demographic variables could be used to predict the likelihood that someone would answer all items on a given scale. The dependent variable was whether respondents answered all items in the scale (yes or no) while the independent variables were the seven demographic characteristics discussed above. The model was not significant for predicting the completion of the DASS-21 depression and stress subscales. However, gender, race, and age were significant for predicting the completion of the DASS-21 anxiety subscale (Table 4).

Logistic regression models with a concordance (c) value below 0.5 are generally considered no better than chance models. Models above 0.7 are considered good while models above 0.8 are strong, with increased odds that the model can correctly predict the dependent variable. The concordance of the DASS-21 anxiety subscale model was 0.811, which is a strong model.

Table 5
Logistic Regression Anxiety = Gender, Race, Age

Predictor	β	SE β	Wald's χ^2	df	p	Odds Ratio (point estimate)
Constant	6.4701	136.4	0.0023	1	0.9622	
Gender			8.7179	2	**0.0128**	
Female	0.4545	0.5616	0.655	1	0.4183	5.975
Male	0.8785	0.5568	2.4892	1	0.1146	9.13
Other	0	.	.		.	
Race			9.1068	3	**0.0279**	
Black or African American	9.8976	409.1	0.0006	1	0.9807	>999.999
Caucasian/white	-2.4888	136.4	0.0003	1	0.9854	6.848
Multiracial	-2.9961	136.4	0.0005	1	0.9825	4.123
SRSC	0	.	.		.	
Age			14.137	5	**0.0148**	
30 or younger	0.1307	0.6629	0.0388	1	0.8438	5.095
31-40	1.4275	0.8943	2.5482	1	0.1104	18.635
41-50	0.3226	0.575	0.3148	1	0.5748	6.173
51-60	-1.0061	0.4221	5.6813	1	0.0171	1.635
61-70	0.6229	0.8795	0.5016	1	0.4788	8.335
71 or older	0	.	.		.	

Test	χ^2	df	p
Overall model evaluation			
Likelihood ratio test	25.688	10	0.0042
Score test	33.8453	10	0.0002
Wald Test	22.1632	10	0.0143
Goodness of fit test			
Hosmer & Lemeshow	3.6529	8	0.887

Note: c = 0.811

In addition, both gender and age were significant in predicting the completion of the AUDIT. However, the concordance was only 0.601, and thus not a very good model.

Table 6
Logistic Regression AUDIT= Gender and Age

Predictor	β	SE β	Wald's χ^2	df	p	Odds Ratio (point estimate)
	-	0.6844	1.1642	1	0.2806	
Constant	0.7385					
Gender			12.9866	2	**0.0015**	
Female	2.0176	0.6264	10.3752	1	0.0013	7.52
Male	1.705	0.6253	7.4347	1	0.0064	5.501
Other	0	.	.		.	
Age			14.2857	5	**0.0139**	
30 or younger	1.0149	0.3754	7.309	1	0.0069	2.759
31-40	0.9683	0.3373	8.2433	1	0.0041	2.634
41-50	0.6834	0.3274	4.3575	1	0.0368	1.981
51-60	0.6665	0.3247	4.2119	1	0.0401	1.947
61-70	0.3041	0.3375	0.8116	1	0.3676	1.355
71 or older	0	.	.		.	
Test			χ^2	**df**	**p**	**Test**
Overall model evaluation						
Likelihood ratio test			26.8567	7	0.0004	
Score test			30.2023	7	<.0001	
Wald Test			27.7764	7	0.0002	
Goodness of fit test						
Hosmer & Lemeshow			4.2548	8	0.8334	

Note: c = 0.601

The only demographic variable that was significant in predicting completion of the DAST-10 was gender. With a concordance of 0.594, this model is not much better than a random chance of predicting the dependent variable.

Table 7
Logistic Regression DAST-10 = Gender

Predictor	β	SE β	Wald's χ^2	df	p	Odds Ratio (point estimate)
Constant	0.9808	0.677	2.0989	1	0.2806	
GenderMFO			14.742	2	**0.0006**	
Female	2.3684	0.7069	11.2244	1	0.0008	10.68
Male	1.7082	0.6939	6.0596	1	0.0138	5.519
Other	0	.	.		.	.
Test			χ^2	**df**	**p**	
Overall model evaluation						
Likelihood ratio test			13.0124	2	0.0015	
Score test			18.2159	2	0.0001	

Wald Test	14.742	2	0.0006	
Goodness of fit test				
Hosmer & Lemeshow	0	1	1	

Note: c = 0.594

Age and whether one had children were significant in predicting the completion of the SCS-SF. With a concordance of 0.641, this model is not much better than a random chance of predicting the dependent variable.

Table 8

Logistic Regression SCS-SF= Age and Whether the respondent had children

Predictor	β	SE β	Wald's χ^2	df	p	Odds Ratio (point estimate)
Constant	1.9235	0.3661	27.5997	1	<.0001	
Age			21.8155	5	**0.0006**	
30 or younger	1.604	0.5311	9.1194	1	0.0025	4.973
31-40	1.7005	0.4803	12.535	1	0.0004	5.477
41-50	1.5521	0.4663	11.0798	1	0.0009	4.721
51-60	0.8872	0.4231	4.3962	1	0.036	2.428
61-70	0.4887	0.4377	1.2467	1	0.2642	1.63
71 or older	0	.	.		.	
Children			4.4241	1	**0.0354**	
No	0.57	0.2671	4.4241	1	0.0354	0.57
Yes		.	.		.	
Test			χ^2	df	p	
Overall model evaluation						
Likelihood ratio test			22.0263	6	0.0012	
Score test			24.4786	6	0.0004	
Wald Test			22.573	6	0.001	
Goodness of fit test						
Hosmer & Lemeshow			1.7441	6	0.9417	

Note: c = 0.641

Neither marital status, time in the field, or profession were significant in predicting completion of any of the scales.

The Chi-Sq results for the individual demographics, in combination with other demographics, show a pattern with many of the demographic variables being statistically significant in many models. However, only the DASS-21 anxiety subscale Logistic Regression model was strong enough to consider that the model was not random.

Discussion

Item non-response is a significant issue in research using summed scales. In this study, 24% of respondents failed to complete all items on all six measures, thus rendering the answers they did answer unusable. Given this impact, more research needs to be conducted into 1)

methods that can increase respondents' completion of all items in summed measures and 2) why respondents neglect to respond to one or more items in a summed scale.

Analysis of missing data, while rarely conducted to this extent, can be used to gain insight into the limitations of generalizing the study results and provide direction for future research. For example, the finding that judges were less likely to respond to all items on the DASS-21 depression subscale, AUDIT, and DAST-10 suggests that caution should be taken in generalizing the findings in the study to judges. Moreover, researchers may want to take additional precautions, and steps need to be taken when surveying judges about sensitive topics such as depression and alcohol and drug use. Judges have often faced significant stress (Maroney, Swenson, Bibelhausen, and Marc, 2023) but more research is needed to fully understand the stress they experience and why they may be reluctant to answer questions that assess their mental health and well-being. Furthermore, additional research should be conducted to examine whether the patterns of missing data are unique to this sample or comparable to other studies of judges and lawyers as well as to other professional groups.

Researchers should be transparent about whether respondents are required to answer all items in a survey, all items in a particular section of the survey, or in a particular measure within the survey, or are not required to answer any items in a survey packet that contains summed scales. Regrettably, the limited research on judges' and attorneys' well-being often fails to provide information about what items must be answered to submit the survey. A 2020 study of 1,034 judges found that 1,026 answered the AUDIT (Swenson, Bibelhausen, Buchanan, Shaheed, & Yetter, 2020). However, the authors do not state whether respondents were required to answer some or all of the items in the survey packet. A subsequent article referencing the Swenson study states that while most respondents completed the AUDIT, they were “permitted to skip any portions of the survey” (Maroney et al., 2023, p. 28). Additional information about whether respondents were required to answer all items in the survey would have been helpful and given additional context for their findings.

The IRB's refusal to let the researchers require that respondents answer questions beyond the informed consent may have prevented a problem known as “straight-lining,” whereby respondents give similar or identical answers to all questions in order to finish the survey (Kim, Dykema, Stevenson, Black, & Moberg, 2019; Mirzaei, Carter, Patanwala, & Schneider, 2002) and, in this case, in order to receive the CLE. While the motivation for answering all items in the different summed scales cannot be known, one hopes and can reasonably assume that the respondents in this study answered the items truthfully, particularly given the precautions taken to protect their confidentiality and privacy.

At the same time, researchers and IRBs should be encouraged to require and perhaps incentivize responses to all demographic questions. Those who failed to respond to demographic items were less likely to answer all questions on many of the summed scales, thus suggesting a relationship between the two.

The results of this study also point to a need for more research on how the order of the measures in a survey affects the completion rate for each measure. The literature on the relationship between the location of scales in surveys and their completion rate is mixed with

some authors suggesting that measures assessing more difficult items be placed later in the survey while others suggest that they should be placed early in the scale when respondents are less prone to respondent fatigue. The DASS-21 was the first scale in the survey; the full 21-item scale and the three subscales had the highest completion rate of any of the measures. Although the 12-item SCS-SF was the last measure in the survey, it did not have the lowest completion rate. Although it has been well documented that respondent fatigue in surveys may lead to item non-response (Lavakras, 2008), this study does not support that concern. Researchers may want to consider putting scales that inquire about sensitive issues at the beginning of the scale. The 10-item AUDIT asks about alcohol use and was placed in the middle of the survey with 21 items asked before it and 22 items asked after it and had the lowest completion rate. Future research should examine how the order of the scales in the survey packet affects their completion rate. For example, studies could be conducted in which the same instruments are included in a survey packet but the order of the instruments varies. This would enable researchers to gain insight into the relationship between the location of the scale in the survey and the completion rate for that scale.

It is further important to examine relationships between measures and patterns of missing data among measures that are both theoretically and statistically related. The AUDIT and DAST-10 both ask about sensitive topics (alcohol and drug use) and there was a moderate relationship between failing to complete these two measures. Weak but significant relationships were found with the AUDIT and SCS-SF. Future research should explore both patterns of missing data among the measures and the nature of these relationships.

Another area of future research should be conducted on law students, attorneys, and judges' reluctance to answer questions about their age, race, and whether they have children. Although respondents answered 94.37% of demographic questions, there was a consistent non-response to these questions when they answered the six summed scales. One explanation is that this population is concerned that these characteristics could be used to identify them and tie them to their responses, despite attempts to limit the identification of respondents in the study.

In sum, missing data in the form of item non-response impacts the results of surveys using summed scales that require all questions in the measure to be answered in order to compute a score on the measure. A clear pattern was found with respondents who failed to answer demographics and also failed to answer all items in the summed scales. Analysis of missing data in surveys using summed scales should examine both missing data in demographic questions and in the summed measures to identify biases in the data caused by patterns of missing data among respondents. This will lead to more confidence in the results and increased awareness about the degree to which the findings may be limited to the specific population who responded or generalizable to broader populations.

References

Anker, J., & Krill, P. R. (2021). Stress, drink, leave: An examination of gender-specific risk factors for mental health problems and attrition among licensed attorneys. *PloS one*, *16*(5), e0250563. https://doi.org/10.1371/journal.pone.0250563

Blair, E., Sudman, S., Bradburn, N. M., & Stocking, C. (1977). How to ask questions about drinking and sex: Response effects in measuring consumer behavior. *Journal of Marketing Research, 14,* 316–321.

Chen, S., and Haziza, D. (2019) Recent Developments in Dealing with Item Non-response in Surveys: A Critical Review. *International Statistical Review, 87,* S192–S218. https://doi.org/10.1111/insr.12305.

de Leeuw, E., Hox, J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics, 19*, 153-176.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44*(2), 227–239. https://doi.org/10.1348/014466505X29657

Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys. *Social Science Computer Review, 37*(2), 214-233. https://doi.org/10.1177/0894439317752406

Krill, P. R., Johnson, R., & Albert, L. (2016). The Prevalence of Substance Use and Other Mental Health Concerns Among American Attorneys. *Journal of Addiction Medicine*, *10*(1), 46–52. https://doi.org/10.1097/ADM.0000000000000182

Lavrakas, P. J. (2008). Respondent fatigue. In *Encyclopedia of Survey Research Methods* (pp. 743-743). SAGE Publications, Inc., https://doi.org/10.4135/9781412963947

Lovibond, S.H., & Lovibond, P.F. (1995). *Manual for the Depression Anxiety Stress Scales*. (2nd. Ed.) Sydney: Psychology Foundation.

Maroney, T., Swenson, D. X., Bibelhausen, J., & Marc, D. (2023). How Are You Holding Up? The State of Judges' Well-Being: A Report on the 2019 National Judicial Stress and Resiliency Survey. *Judicature*, *107*, 22.

McNeeley, S. (2012). Sensitive Issues in Surveys: Reducing Refusals While Increasing Reliability and Quality of Responses to Sensitive Survey Items. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (pp. 377-396). New York, NY: Springer New York.

Mirzaei, A., Carter, S. R., Patanwala, A. E., & Schneider, C. R. (2022). Missing data in surveys: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy, 18*(2), 2308-2316. https://doi.org/10.1016/j.sapharm.2021.03.009

Neff, K. D. (2003). Development and validation of a scale to measure self-compassion. *Self and Identity, 2*, 223-250

Raes, F., Pommier, E., Neff,K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the Self-Compassion Scale. *Clinical Psychology & Psychotherapy, 18*, 250-255.

Rule 10. (2017). *Rules of the Arkansas Judges and Lawyer Assistance Program (JLAP).* https://opinions.arcourts.gov/ark/cr/en/item/1875/index.do#!fragment/zoupio-_Toc97817946/BQCwhgziBcwMYgK4DsDWszIQewE4BUBTADwBdoAvbRABwEtsBaAfX2zgE4B2ADgEYuHACwA2AJQAaZNlKEIARUSFcAT2gBydRIiEwuBIuVrN23fpABlPKQBCagEoBRADKOAagEEAcgGFHE0jAAI2hSdjExIA

Saunders, J.B., Aasland, O.G., Babor, T.F., De La Fuente, J.R. & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction, 88*, 791-804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Skinner H. A. (1982). The drug abuse screening test. *Addictive behaviors, 7*(4), 363–371. https://doi.org/10.1016/0306-4603(82)90005-3

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Sage Publications, Inc. https://doi.org/10.4135/9781412986038

Swenson, D., Bibelhausen, J., Buchanan, B., Shaheed, D., & Yetter, K. (2020). Stress and resiliency in the US Judiciary. *2020 Journal of the Professional Lawyer*, 1.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection, mode, question format, and question context. *Public Opinion Quarterly, 60*, 275–304.

The Role of Artificial Intelligence in Enhancing Scholarly Research – AI Tools Evaluation

Sharon Qi, San Jose State University

Xiaohong Iris Quan, San Jose State University

Taeho Park, San Jose State University

Bobbi Makani, San Jose State University

Abstract

This paper investigates the role of artificial intelligence (AI) in academic research, for example, ChatGPT, Research Rabbit, LitMaps, Scite AI, Elicit, and Copilot. Understanding their impact on research processes is crucial as AI technologies become increasingly sophisticated. This study identifies fifty generative AI software tools and their primary functions for scholars. It examines how these tools differ in functionality complexity, accuracy, reliability, and validity in academic contexts. Through empirical data collected from experiments conducted by the researcher team, the study assesses the effectiveness of fifty AI tools that may potentially assist academic research.

The findings of this research contribute to a deeper understanding of how AI tools can enhance scholarly productivity, streamline research processes, and potentially reshape the future of academic work. By offering practical insights and recommendations, this study aims to inform scholars, educators, and institutions about the opportunities and challenges associated with incorporating AI into academic research.

Keywords: Artificial intelligence (AI), AI tools evaluation, AI in scholarship, Research process efficiency

Introduction

Artificial Intelligence (AI) has started to revolutionize various fields in academia, where AI-powered tools are increasingly utilized to enhance research productivity and writing efficiency. According to the Economist report, 10% or more of abstracts for papers in scientific journals now appear to be written at least in part by large language models; In fields such as computer science, that figure rises to 20% (Economist, 2024). According to another new study by Harvard Business School, when AI is used by highly skilled workers, it can improve a worker's performance by as much as 40% compared with workers who don't use it (Somers, 2023). AI has recently been defined as "the use of computational machinery to emulate capabilities inherent in humans, such as doing physical or mechanical tasks, thinking, and feeling" (Huang and Rust, 2021, p.31). More and more AI tools have emerged for academic use. The application of AI tools in academia ranges from literature reviews, data analysis, to abstracts compiling and writing aids, such as grammar check and reference management. The promise of AI goes beyond simple automation. It also includes the potential to extract patterns and insights from large volumes of data that are sometimes invisible to human researchers and can even synthesize new research ideas, allowing researchers to focus more on critical thinking and sophisticated problem-solving.

The academic community has shown a keen interest in exploring the effectiveness of specific AI tools, especially since the release of ChatGPT in November 2022, with the primary focus predominantly being on individual tools like ChatGPT (Aydın, 2023; Dwivedi, Kshetri, Slade, Jeyaraj, and Kar, 2023; Hosseini, Rasmussen, and Resnik, 2023; Nguyen-Trung, Saeri, and Kaufman, 2023; Salvagno, Taccone, and Gerli, 2023). But there are a few exceptions with focus on other AI tools. For instance, Kurniati and Fithriani (2022) examined the use of Quillbot for enhancing English academic writing, and Marzuki, Widiati, Rusdin, Darwin, and Indrawati (2023) investigated a few AI writing tools on student writing quality from the perspective of English as a Foreign Language (EFL) teachers and demonstrated the benefits of integrating some AI writing tools for EFL students. However, there remains a significant gap in the literature concerning a comprehensive and systematic evaluation of the wide variety of AI tools available for academic work.

This evaluation study aims to address this gap by providing a comprehensive evaluation of multiple AI tools designed for academic use. By identifying and assessing the capabilities, benefits, and limitations of approximately fifty AI tools, this study seeks to offer valuable insights into how these tools can enhance the academic research and writing process. Such an evaluation is crucial for academics to make informed decisions about how to integrate AI technologies into their research work, ultimately advancing scholarly productivity and innovation.

Literature Review

The scope of the literature review in this section focuses on studies on AI software tools for scholars, ranging from AI tools that assist literature review, academic writing, and publication review, to other writing-related assistance such as reference management, as well as concerns about the reliability of AI tool, such as ChatGPT, and the related ethics.

AI assisting literature review

For scholars to conduct research, a literature review usually serves as one of the initial steps. AI's potential to expedite literature reviews has been explored in several studies, although the use of AI in this context is in an early stage of development.
Many studies have highlighted the potential positive role of AI in the research process, although concerns about reliability have been repeatedly raised. Wagner, Lukyanenko, and Paré (2021) provided a comprehensive overview of AI's capabilities in this area, which reviewed how the reported AI applications prior the release of ChatGPT were applied to a set of steps of the literature review process, for example, steps of problem formulation, literature review, searching, screening for inclusion, quality assessment (i.e., methodological flaws and source of bias), data extractions (i.e., getting relevant information), and data analysis & interpretation (i.e., either descriptive syntheses or inductive work such as theory generation), as well as proposing a research agenda for AI-based literature reviews (AILRs) at three different levels (i.e., supporting infrastructure, methods and tools, and research practice). They highlighted AI's efficiency in handling large volumes of documents and facilitating literature synthesis. Johnson, Bauer, and Niederman (2021) mentioned a tool called ORA, which is a dynamic network analysis tool for analyzing Scopus or other online bibliographic sources (e.g., Ebsco, Google Scholar, Orcid, ProQuest, Web of Science). Unlike two studies reviewed here that were published pre-Chat GPT age, Nguyen-Trung, Saeri, and Kaufman (2023) demonstrated how AI tools like ChatGPT and ChatPDF could enhance evidence reviews, despite limitations such as inconsistent results and certain errors.

Although these AI-powered tools demonstrated some comprehension of research concepts, it has been repeatedly reported that the AI tools sometimes misinterpreted material or generated misleading descriptions or summaries of those concepts. For example, in the test of using ChatGPT for literature review in the field of medicine, Haman and Školník (2023) found that in only 17 out of 50 instances, the articles could be located within the databases (such as Google Scholar, PubMed, Semantic Scholar), while 66% of the references produced by ChatGPT were non-existent papers.

At the methodological level, Guler, Kirshner, and Vidgen (2023) highlighted the effectiveness of combining machine learning and ChatGPT in literature reviews, noting improvements that have made in identifying research topics and generating content. While they used machine learning to identify research topics, ChatGPT was used to assist in labeling the topics, generating content, and improving the efficiency of academic writing.

AI assisting academic writing

AI has the potential not just to improve both the efficiency of literature reviews but also the overall productivity of academic writing. Significant exploration has been conducted into how certain AI tools can be effectively leveraged to enhance the academic writing process, as well as to understand their limitations.

Abdul, Mathew, Ahmad Saad, and Alqahtani, (2021) provided a comprehensive review of AI's role in scholarly writing. They identified tools across various categories, including literature

search and review, writing and editing, references and citations, review and workflow, plagiarism checking, and journal selection. However, since their paper was published in 2021, before the release of ChatGPT, like some other reviewed papers, the tools they identified do not encompass the recent advancements in AI within this field.

Khalifa and Albadawy (2024) conducted another comprehensive review study on AI-assisted academic writing. After an initial screening of 217 studies, they filtered their selection further down to 24 studies. Through analysis, they identified six core domains where AI supports academic writing, including 1) idea development and research design, 2) content development and structuring, 3) literature review and synthesis, 4) data management and analysis, 5) editing, review, and publishing support, and 6) communication, outreach, and ethical compliance. They summarized each of these 24 papers by their titles, main focus, key findings, what the AI application is about, limitations, and recommendations. After they mapped these 24 studies with the six domains, they found that domain 1) was the least researched, while domains 5) and 6) were mostly researched. In their conclusion, they addressed the importance of integrating AI tools into the academic research process, while they mentioned the ethical and transparent use of AI tools, as well as the need of training in using AI tools. They urged people to continue researching AI's impact on academic research since this is an ever-evolving process. They also made a few recommendations for AI technology in assisting research including "developing advanced AI tools for hypothesis formulation and predictive analysis", and "establishing ethical frameworks for AI use" (p10).

Most studies on AI for scholars in the current literature have predominantly focused on ChatGPT, an AI tool known for its comprehensive features, particularly in academic writing. A notable study in this area is the opinion research conducted by Dwivedi, Kshetri, Slade, Jeyaraj, and Kar (2023). The study involved many authors. They extensively investigated the impact of ChatGPT through 43 contributors across fields including computer science, marketing, information systems, education, policy, hospitality and tourism, management, publishing, and nursing. The evolvement of AI technologies was overviewed, and the positive impact of ChatGPT on various industries was acknowledged. The authors raised concerns such as "threats to privacy and security, and consequences of biases" (p3). They suggested further research in three areas along with many detailed research directions: 1) "knowledge, transparency, and ethics"; 2) "digital transformation of organizations and societies"; and 3) "teaching, learning, and scholarly research." (p3)

Salvagno, Taccone, and Gerli (2023) discussed the utility of AI chatbots like ChatGPT in scientific writing, assisting researchers and scientists in organizing material, generating an initial draft, and/or in proofreading. They evaluated tools such as Elicit and compared the performance of AI chatbots with that of humans. They concluded that while AI tools were useful, they did not surpass humans in highly technical areas, particularly in selecting the most appropriate wording. They also emphasized that human oversight was crucial to ensure accuracy and prevent plagiarism. This aligned with Haman and Školník's (2023) findings on the limitations of AI tools in generating accurate academic references.

Ray (2024) highlighted the potential of ChatGPT in early career research scholarship, mentioning its role in refining hypotheses and conducting literature reviews. However, the discussion should have delved deeper into these aspects.

Marzuki, Widiati, Rusdin, Darwin, and Indrawati (2023) conducted a qualitative study examining the influence of AI writing tools on student writing. He interviewed four English as Foreign Language (EFL) teachers who had years of EFL teaching experience and a certain time of using AI tools as part of their teaching curriculum. Based on his research review on AI's impact on teaching writing, he designed an interview instrument and collected opinions and feedback from these participants. Each teacher used 3-4 different AI tools, including Quillbot, WordTune, ChatGPT, Essay Writer, PaperPal, and Jenni where QuillBot and WordTune were used by four different teachers. The overall impression was that these AI tools improved the content and organization of student writing, although the level of potential positiveness of AI tools to students' English writing skills was different. This qualitative study provided a rich insight into EFL teachers' teaching experiences using AI tools.

AI impact on the publication industry

In addition, AI-supported technologies are rapidly changing the publishing industry including reviewing and editing, which is another key domain integral to the research process. For instance, AI-driven software like ChatGPT, Grammarly, and Paperpal can correct grammatical errors and improve writing style. Tools like Zotero, Mendeley, and EndNote are indispensable for literature management. Turnitin and Copyscape stand out in the domain of plagiarism detection, employing extensive databases to verify the originality of academic works. The UK Publishers Association investigated the role of AI in the publishing industry, including AI investments and the obstacles faced by the sector. Their findings were published in a report titled *The Role of Artificial Intelligence in Publishing* (2020), which revealed that most publishing sectors believe AI "will be important over the next five years" (p3). The report also noted that "AI investment in the sector has just begun," with "larger publishers leading the drive" (p3).

AI-assisted writing education

Researchers have also explored the impact of AI on students and the educational process. A study of 343 communication instructors revealed a collective view that AI-assisted writing is widely adopted in the workplace and requires significant changes to instruction despite challenges such as less critical thinking and authenticity in writing (Cardon et al, 2023). For instance, Cribben and Zeinali (2023) reviewed the benefits and limitations of ChatGPT in business education and research, noting its utility in designing courses, creating content, and grading. Steele (2023) argued that AI tools like ChatGPT could empower students by enhancing their comprehension, research, and composition skills. Nevertheless, he also noted the AI-related threats to traditional education systems, such as measurement problems and skill devaluation. Mishra (2024) investigated the integration and experiences of academic professionals with AI tools in their pedagogical practices and illustrated a broad understanding and adoption of AI tools. The findings highlighted the need for adequate training and ethical guidelines for responsible AI use.

Kurniati and Fithriani (2022) studied how post-graduate students viewed Quillbot as a digital tool for English academic writing by employing a qualitative case study design involving 20 post-graduate students majoring in English education. The findings revealed that the postgraduate students in the study responded positively to using Quillbot to assist them in improving the quality of their writing.

Reliability and ethical concerns

Mandai, Tan, Padhi, and Pang (2024) highlighted the propensity of AI tools to generate content based on statistical probabilities rather than understanding, leading to errors and hallucinations. Dashti, Londono, Ghasemi, and Moghaddasi (2023) tested whether ChatGPT could find same articles in *The Journal of Prosthetic Dentistry (JPD)* as people searched directly in the journal and Google Scholar using a set of same keywords at different time. They found the results did not match the papers that ChatGPT had generated. Furthermore, all 75 articles provided by ChatGPT were not accurately located in the JPD or Google Scholar databases.

In an editorial for *the Journal of Accountability in Research*, researchers (Hosseini, Rasmussen, and Resnik, 2023) explored the performance of ChatGPT in writing tasks. Their findings revealed that the chatbot sometimes produced responses that were either entirely incorrect or not pertinent to the given topics. Consequently, they advocated for a rigorous review process, recommending that "any section of a manuscript created by an NLP system should be meticulously examined by a domain expert to ensure its accuracy, detect any biases, maintain relevance, and evaluate the reasoning presented (p1)."

Ciaccio (2023) addresses the necessity of recognizing AI's role in scientific writing. Basic AI assistance, such as spell-checking and grammar correction, typically does not require acknowledgment and can be effectively managed by AI tools. However, more extensive editorial interventions, including content editing—such as reorganizing paragraphs or sections, rewriting passages, and adding or deleting content—should be transparently disclosed by authors when submitting manuscripts for publication.

Aydın (2023) explored the use of Google Bard for generating literature reviews, comparing its performance to ChatGPT. They found both tools showed promise but exhibited higher plagiarism rates and occasional inaccuracies, underscoring the need for careful monitoring and ethical guidelines.

According to research by Casal and Kessler (2023) on *Applied Linguistics journals*, the journal reviewers were largely unsuccessful in identifying AI versus human writing, and many editors believed there are ethical uses of AI tools for facilitating research processes. Further research is needed to address the matter.

Research Questions

In summary, the literature as reviewed above presented both the potential benefits and the challenges of AI and AI-based tools in academia. These tools can potentially enhance research productivity and writing quality significantly, although concerns of inaccuracy and ethical warnings were consistently addressed. We found that there was very little systematic and in-depth research conducted regarding AI-assisted academic tools' quality evaluation, such as

accuracy, reliability, and scalability, as well as usability. This is a critical issue as we embrace AI in the academic research process, and it warrants careful consideration. To address this gap and help provide comprehensive guidelines in using AI-assisted academic tools, we conducted a study with the following primary goals and research questions. We present the results in the later part of this explorative paper.

The primary goal of this paper is to explore the emergent AI tools that can potentially assist scholars in doing their academic research more effectively and efficiently, using a thorough evaluation methodology that supports them in making a sound decision on what AI tool they should use for different purposes.

***Research Question 1*:** What generative AI software tools are currently available for academic scholars, and how do they differ in their primary functions and complexity?

***Research Question 2*:** How do these AI tools vary in terms of accuracy, reliability, validity, usability, and scalability in assisting scholarly work, as well as the associated costs?

***Research Question 3*:** What are the key recommendations in helping researchers decide on which AI tool may fit what needs, and along what concerns regarding the adoption and integration of AI tools in academic research?

Methodologies

Participants

A student research team of eight, who majored in software engineering or computer science, six as undergraduates and two graduates, supervised by the first two authors, participated in data collection. The team was sponsored by school faculty research funding and school early career training funding. Weekly meetings were held for coaching, data collection review, data cleaning, and data organizing.

Data collection procedure

We processed the data collection by three steps below:

Step 1: We started a pilot testing with one AI tool for 3 weeks. Based on the pilot experiences, we figured out an effective way to collect data.

Step 2: Once we created a best practice guideline and defined the scope of exploration, we moved on to exploring a large set of AI-assisted research tools. We investigated these tools from the research-related functions' perspective, such as searching and finding references, processing literature review, judging reference's relevance and references/citations' mapping, helping paper writing idea creating and structuring, assisting wording selection, and proofreading.

Step 3: We further filtered and shortened the tool list, and re-evaluated them regarding accuracy, reliability, and validity. Then, we explored the complexity and usability of the tools. The ranking was based on a rigid rating criterion. See the details below. The data was processed using a cross-tester rating, and the average was calculated to mitigate the subjectivity.

The data collection process started in November 2023 and ended in June 2024.

Initial data sources

To systematically identify and select AI tools relevant for scholarly use, we employed a comprehensive approach that involved multiple diverse arrays of reputable data sources. The initial phase of our data collection involved sourcing potential AI tools from websites dedicated to technology and education, online rankings of AI tools, peer-reviewed academic papers, expert blogs, and credible news outlets. The selection criteria were rigorously defined to ensure that only legitimate and well-regarded tools were considered. Specifically, tools that had been consistently highlighted in the literature for their innovative capabilities, widespread usage, or endorsements by experts in the field were given precedence. We also cross-referenced these tools with online reviews and user feedback to gauge their effectiveness and reliability in academic settings. A list of over 80 AI tools were initially identified.

Data sources filtering

This initial list was subsequently refined through a careful selection process that aimed to ensure a balanced representation of AI tools catering to different aspects of scholarly work. The final selection comprised 50 AI tools, which were chosen based on their ability to perform key functions critical for academic productivity. These functions include searching for papers on specific topics, reading and analyzing academic papers, providing summaries, checking searched sources' relevance, reference mapping, citation assistance, content organizing, and assisting with academic writing. The selected tools span a range of capabilities from advanced search engines that utilize AI to pinpoint relevant academic papers, to sophisticated text analysis tools that offer in-depth comprehension aids and writing assistants that support the drafting and editing process. The diversity of the selected tools ensures that they collectively address the broad spectrum of tasks that scholars typically engage in during their research and writing processes.

Cross-tester rating

These filtered 50 AI tools were ranked based on specifically given criteria. See the next section for details. Each selected tool was evaluated by two testing team members (evaluation participants). Their rating average was calculated to mitigate subjectivity. See the final matrix in the research findings section later in this paper.

This evaluative process was designed to capture the user experience and practical utility of each tool from the perspective of end-users, namely, how each participant would most benefit from these technologies. By incorporating user feedback into our methodology, we aimed to provide a holistic assessment of the AI tools, balancing technical functionality with practical usability. The resulting rankings offered valuable insights into which tools were most effective in supporting academic scholarship and which features were mostly valued by users in an academic context. This methodological approach, grounded in both extensive literature review and empirical user feedback, ensured that the selected AI tools are not only technically proficient but also practically relevant to the needs of scholars. The careful selection process, combined with rigorous evaluation by the research team, provided a robust foundation for understanding the

current landscape of AI tools available for academic purposes and their potential impact on scholarly productivity.

Rating factors/ranking scale

We outlined a detailed protocol used for our research team to evaluate the filtered 50 AI tools across a range of critical metrics. The selected tools were rigorously tested based on eight key criteria: cost, accuracy, reliability, validity, function complexity, usability, efficiency, and scalability. The criteria guided a comprehensive assessment of each tool's performance and explored not only their effectiveness but also their adaptability to the diverse needs of academic work. See below for details.

Cost. The cost of using each AI tool was a consideration in our evaluation process. We examined both upfront costs and ongoing expenses, including subscription fees and any additional charges associated with premium features or updates. The affordability of each tool was assessed relative to its functionality, to determine whether the cost is justified by the tool's capabilities and benefits. We did not use a Likert scale for the cost factor but provided direct cost information based on three categories that we identified through the research. The three categories include free, freemium, or fee-based. The cost information would be able to accommodate the budgets of researchers, faculties, students, and institutions who need to plan their budgets in adopting them.

Accuracy, reliability, and validity. Accuracy is a crucial metric for AI tools, particularly in academic settings. We tested if the tool did exactly what it said (e.g., answer the questions correctly). We assessed the accuracy of each tool in performing specific tasks such as grammar checking, data extraction, and content generation. This involved testing the tools against a set of benchmark tasks where the expected outcomes were well-defined. For example, grammar-checking tools were evaluated based on their ability to identify and correct syntactical errors, while data extraction tools were tested on their precision in retrieving relevant information from complex datasets. The tools were scored based on their error rates and the relevance of the results they produced. We adopted the Likert scale of one through five for the ranking purpose. Reliability was measured by examining the consistency of each tool's performance over multiple use cases across different contexts. We tested if the tool performed the same function across three different instances (i.e., three different academic papers from the literature list provided). This involved repeated testing of the tools under varying conditions to ensure that they could deliver consistent results irrespective of the complexity of the task or the volume of data processed. A tool's reliability was determined by its consistency and ability to function without failure or significant performance degradation over time. Tools that demonstrated high stability and consistent output were ranked higher. Like reliability, reliability is essential to tools that scholars depend on for critical academic tasks.

The validity of each AI tool was assessed by determining the extent to which it measures what it claims to measure. We tested whether what was done was true, for example, whether the generated citation was truly existing. This involved comparing the tool's outputs with established standards or expert assessments to ensure that the tool accurately fulfilled its intended purpose. A tool's validity was considered high if its results closely matched the expected outcomes or expert judgments, indicating that the tool was effective in achieving its stated objectives.

Likert Scale	1) answer your questions exactly as expected;	2) complete the task with consistent quality across three or more different instances;	3) each verification proves the truth of the generated content (e.g., the citation is truly existing, or the data facts are true).
5	100%	100%	100%
4	80%	80%	80%
3	50%	50%	50%
2	30%	30%	30%
1	0%	0%	0%

Table 1: Likert Scale for Accuracy, Reliability, and Validity

Function Complexity. Function complexity refers to the range and sophistication of the functions provided by each AI tool. Tools were evaluated based on their ability to perform a diverse array of tasks, as well as the depth of customization and control they offered to users. We measured the complexity of a tool based on the basic scholarly functions, such as literature searching, relevance checking, paper reading (reviewing and analyzing), reference and citation, and paper writing.

Likert Scale	Function/Features	Criteria Description
5	Functions: Searching Reading Writing Features: 5-8 as listed on the criteria description	Functions: All 3 functions searching, reading, and writing are present Features: 5-8 or more features (e.g., finding paper, checking relevance, summary, detailed paper analysis, editing, chatting, citation, references, related paper finding, recommending the writing outline, etc.). Part of a suite of tools: The tool is linked to another tool, not isolated, as a part of a suite, e.g., ChatPDF vs ChatGPT. Each

Likert Scale	Function/Features	Criteria Description
	Part of a suite of tools? Yes	claimed feature is effective, and the database is also comprehensive.
4	Functions: 1-2 functions of those listed above Features: 1-4 as listed on the criteria description	Functions: 1-2 functions (only reading and writing, or just searching, with a limited database or a special database) Features: 1-4 features (editing, citation, etc.) provided very thoroughly. Part of a suite of tools: The tool is relatively isolated or stands alone, but it performs one function/feature effectively. The effectiveness exceeds the tool with the same function/feature in the more comprehensive tool suite that is scaled 5
3	Functions: 1-2 functions of those listed above Features: 1 as listed on the right, but unique	Functions: 1-2 functions (either reading or writing, or both, but no searching function at all) Features: Some limited feature(s) provided effectively, e.g., summarizing a paper only, or checking relevance only. But it is so unique that it cannot be substituted.

Likert Scale	Function/Features	Criteria Description
2	Functions: 1-2 functions of those listed above Features: 1 as listed on the right, not unique	Functions: 1-2 functions and provides some limited features that can be easily substituted by other tools.
1	Functions: 1 function of those listed above But it is very difficult to duplicate the result due to its technical roadblocks	Functions: 1 function, and cannot perform reliably, e.g., it cannot be opened, or it requires a long time of waiting for authorization.

Table 2: Likert Scale for Function Complexity

Usability. Usability was assessed by evaluating the ease of use and the user interface (UI) design of each tool. This metric focused on how intuitive and accessible the tool was for researchers who had various levels of technical expertise background. We analyzed the overall user experience. Tools that provided a seamless and user-friendly experience, with clear navigation and minimal learning curve, were rated higher in usability.

Likert Scale	Time needed	Criteria Description
5	<1 min	Users can figure it out how to use the tool in a few seconds (<1 min) and find buttons for different features in 1-5 minutes.
4	10-15 min	The tool has a 10 to 15-minute learning curve
3	16 -30 min	The tool requires taking a tutorial training to learn it for 30 minutes or more
2	>30 min	The tool is very hard to use, and no user guides are provided at all
1	> 1 hour	The tool simply doesn't work, after an hour of engagement with it

Table 3: Likert Scale for Usability

Efficiency and scalability. Efficiency was measured by the time and resource consumption of each tool in performing tasks. We evaluated how quickly each tool could complete assigned tasks compared with the traditional manual way. Tools that performed tasks swiftly were ranked higher, as efficiency is critical in academic settings where time may be limited. Additionally, we considered how well the tools managed large datasets and complex calculations, with no compromising the accuracy and reliability.

Scalability refers to a tool's ability to handle increasing amounts of work or data without compromising performance. Tools that demonstrated the ability to scale effectively, handling higher volumes of data or more complex tasks without significant loss of performance, were rated higher. Scalability is particularly important for tools used in research environments where the scope of work may expand over time. A Likert scale was used and calculated with a percentage, which was the recorded time processed with the AI tool divided by the recorded time processed by transitional manual way to reach a rating from one to five. The unit of time recorded was in seconds.

Likert Scale	**Criteria in %**	**Criteria Description**
5	25-39% or below	A Likert scale was used and calculated with a percentage, which was the recorded time processed with the AI tool divided by the recorded time processed by transitional manual way to reach a rating from 1 to 5. The unit of time recorded was seconds. If it is 5, it means that they are very different, and the AI tool is dramatically faster and more efficient across multiple testing instances.
4	40-54%	It means that they are quite different, and the AI tool is more than twice more efficient than the traditional tool.
3	55-69%	It means that they are different, and the AI tool is faster, but not more than twice as efficient as the traditional tool.
2	70-84%	It means that they are somewhat different, and the AI tool is slightly better
1	85-100%	It means that they are somewhat similar.

Table 4: Likert Scale for Efficiency and Scalability

Findings

Figure 1 showed how the filtered 50 AI tools that we investigated were classified by these functions:

- Read (9 in total),
- Read & Write (8 in total)
- Search, Map, & Citation (10 in total)
- Search & Read (2 in total)
- Search, Read & Write (8 in total)
- Write (13 in total)

Based this classification, we built a Table 5 series, in which each category belonged to one function, for example, Table 5(a) for Read, Table 5(b) for Ready & Write, Table 5(c) for Search, Map (i.e., Mapping), & Citation, Table 5(d) for Search & Read, Table 5(e) for Search, Read & Write, and Table 5(f) for Write.

At the bottom of each subtable (i.e.,Table 5 series, or Table 5(a), Table 5(b), Table 5 (c), Table 5 (d), Table 5 (e), and Table 5 (f), you will see sub-total evaluation scores by these indexes:

- Accuracy
- Reliability
- Validity
- Complexity
- Usability,
- Efficiency and Scalability
- Total score

From the individual tool's perspective, we aggregated evaluation scores in the very far right side column of each sub-table 5 series (i.e., Tables 5(a) - Table 5 (e)) into Table 5(f). Here, 32 out of 50 or 64% of the tools were scored at 24-29 (here 30 as the total scores). Among them, 14 of them, or 28% of the tools were scored at 27-29 (out of 30). Three of the tools that notably stood out and nearly achieved full scores were ChatPDF, Copilot, and BingAI, due to their higher scores in the factor of complexity. This data indicates that AI tools that have rich features and good complexity is more favored.

Table 5(f), as a summary table, indicated that 39 out of 50 or 78% were scored at four or above for **accuracy**, 41 out of 50 or 82% at four or above for **reliability**, 42 out of 50 or 84% at four or above for **validity**, 27/50 or 54% at four or above for **complexity**, 38 out of 50 or 76% at four or above for **usability**, and 32 out of 50 or 64% at four or above for **efficiency and scalability**. This data indicates that more AI tools performed stronger in reliability and validity, with the accuracy seemingly acceptable overall, but fewer AI tools were able to provide comprehensive features.

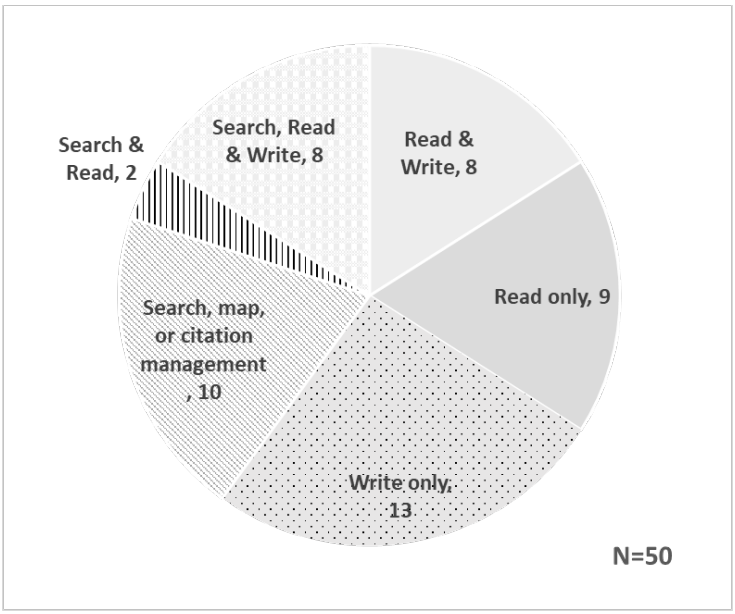

Figure 1: AI Tools for Scholars by Function Types

Table 5 (a)									
AI Tool Name (N=50)	**Major Function (s)**	**Cost**	**Accuracy**	**Reliability**	**Validity**	**Complexity**	**Usability**	**Efficiency & Scalability**	**Total Scores by each tool**
Mindgrasp AI	Read	Fee-based	3	3	3	3.5	4.5	4.5	21.5
Citation.ai	Read	Free	4	3.5	5	2.5	5	4.5	24.5
Summarize Paper	Read	Free	4.5	4.5	4.5	2.5	3.5	4.5	24
Resoomer	Read	Free	3	3	4	3	3.5	4	20.5
PDF AI	Read	Free	4.5	4.5	4.5	3.5	4.5	4.5	26
Myreader	Read	Freemium	3.5	4	4.5	3	4.5	4.5	24
Docalysis	Read	Freemium	4.5	4.5	4.5	4	4.5	4.5	26.5
SciSummary	Read	Freemium	3.5	3.5	3.5	3	4	4.5	22
Typeset	Read	Freemium	3.5	4.5	4.5	3.5	4.5	4.5	25

Table 5 (a): 50 AI Tools for Scholars by Function Types

Table 5 (b)									
AI Tool Name (N=50)	**Major Function (s)**	**Cost**	**Accuracy**	**Reliability**	**Validity**	**Complexity**	**Usability**	**Efficiency & Scalability**	**Total Scores by each tool**
Scholarcy	Read, Write	Fee-based	4	4	4	4	4	4	24
PDFgearCopilot	Read, Write	Free	5	5	5	4	5	3	27
Chatgpt	Read, Write	Freemium	4.5	5	3.5	4.5	5	5	27.5
ChatPDF	Read, Write	Freemium	5	5	5	4.5	5	5	29.5
Sider AI	Read, Write	Freemium	3.5	3.5	3	5	3.5	4.5	23

Scispace	Read, Write	Freemium	4	4	3	5	4	5	25
ScholarAI	Read, Write	Freemium	5	4.5	5	4	4	4	26.5
Bit.ai	Read, Write	Freemium	4	5	5	4	5	4	27

Table 5 (b): 50 AI Tools for Scholars by Function Types

Table 5 (c)									
AI Tool Name (N=50)	**Major Function (s)**	**Cost**	**Accuracy**	**Reliability**	**Validity**	**Complexity**	**Usability**	**Efficiency & Scalability**	**Total Scores by each tool**
EndNote	Search, Map, or Citation	Fee-based	5	5	5	2	3	3	23
zotero	Search, Map, or Citation	Free	4.5	5	4.5	3	2.5	1.5	21
Readcube	Search, Map, or Citation	Fee-based	5	5	4	4	4	1	23
Semantic Scholar	Search, Map, or Citation	Free	5	5	5	1	5	2	23
Mendeley	Search, Map, or Citation	Free	5	5	5	2	4.5	2.5	24
SourceData	Search, Map, or Citation	Free	1	1	1	1	2	1	7
Connected Papers	Search, Map, or Citation	Freemium	5	5	5	3.5	4	2	24.5
Research Rabbit	Search, Map, or Citation	Free	5	5	5	4	3.5	5	27.5
Citation Gecko	Search, Map, or Citation	Free	4.5	4.5	4	2.5	2.5	2	20
Litmaps	Search, Map, or Citation	Freemium	5	5	5	4	4	5	28

Table 5 (c): 50 AI Tools for Scholars by Function Types

Table 5 (d)									
AI Tool Name (N=50)	**Major Function (s)**	**Cost**	**Accuracy**	**Reliability**	**Validity**	**Complexity**	**Usability**	**Efficiency & Scalability**	**Total Scores by each tool**
Semantic Reader	Search, Read	Free	4.5	4.5	4	3	3.5	3	22.5
Scinapse	Search, Read	Freemium	3.5	4	4	2	5	3	21.5
Scite AI	Search, Read, Write	Fee-based	5	4.5	5	4	4	5	27.5
Elicit	Search, Read, Write	Fee-based	5	4.5	5	4.5	4	5	28
Meta AI	Search, Read, Write	Free	4	4.5	4.5	4.5	4.5	4.5	26.5
Consensus	Search, Read, Write	Freemium	5	5	5	4.5	4.5	4.5	28.5
Copilot Sidebar	Search, Read, Write	Freemium	5	5	5	5	5	5	30
Bing AI (Bing Copilot)	Search, Read, Write	Freemium	5	4.5	4.5	5	5	5	29
Gemini	Search, Read, Write	Freemium	4.5	4	4	5	4.5	4	26
Claude	Search, Read, Write	Freemium	4.5	4.5	4.5	4.5	4.5	3.5	26

Table 5 (d): 50 AI Tools for Scholars by Function Types

Table 5 (e)									
AI Tool Name (N=50)	**Major Function (s)**	**Cost**	**Accuracy**	**Reliability**	**Validity**	**Complexity**	**Usability**	**Efficiency & Scalability**	**Total Scores by each tool**
Word AI	Write	Fee-based	3.5	4.5	4.5	3	4.5	4.5	24.5
AI Writer	Write	Fee-based	4.5	3.5	4.5	4.5	5	5	27
ProDream Inc.	Write	Freemium	4.5	5	5	4	5	4.5	28
Lightkey	Write	Freemium	1	1	1	1	1	1	6
Grammarly	Write	Freemium	4.5	5	4	3.5	5	4.5	26.5
Scinote	Write	Freemium	3	3	3	4	2	3	18
Trinka	Write	Freemium	4.5	4	4	3.5	4.5	4	24.5
Crimson.ai	Write	Freemium	4	4	3	3	4	3	21
Junia	Write	Freemium	4.5	4.5	4	4	5	4.5	26.5
writesonic	Write	Freemium	4.5	4.5	4.5	4	5	4.5	27
Quillbot	Write	Freemium	5	5	5	4	5	5	29
jenni.ai	Write	Freemium	4	4.5	5	3	4.5	2.5	23.5
paperpal	Write	Freemium	5	5	5	4.5	3.5	2	25

Table 5 (e): 50 AI Tools for Scholars by Function Types

Table 5 (f)									
AI Tool Name (N=50)	**Major Function (s)**	**Cost**	**Accuracy**	**Reliability**	**Validity**	**Complexity**	**Usability**	**Efficiency & Scalability**	***Total Scores by each tool***
	All function types	**Total Scores by Each Criteria**	**39**	**41**	**42**	**27**	**38**	**32**	*32, 64% /50 tools (Scored 24-29)*
		Total % out of 50	78%	82%	84%	54%	76%	64%	*14, 28% /50 tools (Scored 27-29)*

Table 5 (f): 50 AI Tools for Scholars by Function Types

Looking into the price models, among the same filtered 50 tools that we evaluated (see Figure 2), 29 out of 50 or 58% were freemium (i.e., free at the beginning, following up with a fee charge), 13 out of 50 or 26% free, and 8 out of 50 or 16% fee-based, normally ranging from $10 per month to $40 per month. Such data indicates that the AI tools that have the writing feature are more likely to have a fee-charging. See Figure 2 below for more details.

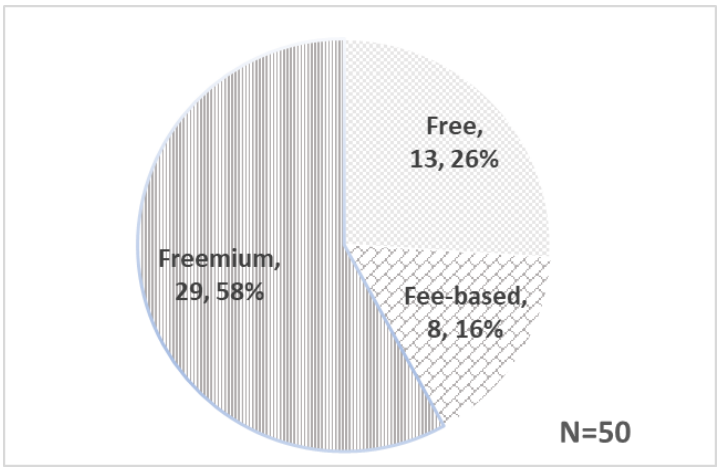

Figure 2: AI Tools for Scholars by Pricing Models

If we re-classified the same 50 AI tools by their price models (see Table 6), nine of the ***read-only*** type of AI tools had no significant differences in their accuracy, reliability, validity, complexity, usability, and efficiency/scalability. The eight ***read-and-write*** types of AI tools and those of the free price model appeared weaker in their accuracy, reliability, validity, complexity, usability, and efficiency/scalability than the other two models, particularly those with the freemium model. For ten of the ***searches, map, or citation*** type of AI tools, as well as two of ***search and read*** type of AI tools, they lose points for not having function complexity. For eight of the ***searches, read and write*** the type of AI tools, along with 13 of the ***write-only*** type of AI tools, they all scored relatively high, where the AI tools of the fee-based model stood out. This data indicates that the AI tools are still having challenges in improving their quality in reading and writing, while the AI tools may offer higher quality assistance if users pay for them.

AI Tools by Function Types and Fee Models	**Accuracy**	**Reliability**	**Validity**	**Function Complexity**	**Usability**	**Efficiency & Scalability**
Read	**3.78**	**3.89**	**4.22**	**3.17**	**4.28**	**4.44**
Fee-based (1)	3	3	3	3.5	4.5	4.5
Free (4)	4	3.88	4.5	2.88	4.13	4.38
Freemium (4)	3.75	4.13	4.25	3.38	4.38	4.5
Read, Write (8)	**4.38**	**4.5**	**4.19**	**4.38**	**4.44**	**4.31**
Fee-based (1)	4	4	4	4	4	4

Free (1)	5	5	5	4	5	3
Freemium (6)	4.33	4.5	4.08	4.5	4.42	4.58
Search, Map,or Citation (10)	**4.5**	**4.55**	**4.35**	**2.7**	**3.5**	**2.5**
Fee-based (2)	5	5	4.5	3	3.5	2
Free (6)	4.17	4.25	4.08	2.25	3.33	2.33
Freemium (2)	5	5	5	3.75	4	3.5
Search, Read (2)	**4.00**	**4.25**	**4.00**	**2.50**	**4.25**	**3.00**
Free (1)	4.5	4.5	4	3	3.5	3
Freemium (1)	3.5	4	4	2	5	3
Search, Read, Write (8)	**4.75**	**4.56**	**4.69**	**4.63**	**4.5**	**4.56**
Fee-based (2)	5	4.5	5	4.25	4	5
Free (1)	4	4.5	4.5	4.5	4.5	4.5
Freemium (5)	4.8	4.6	4.6	4.8	4.7	4.4
Write (13)	**4.04**	**4.12**	**4.04**	**3.54**	**4.15**	**3.69**
Fee-based (2)	4	4	4.5	3.75	4.75	4.75
Freemium (11)	4.05	4.14	3.95	3.5	4.05	3.5

Table 6: 50 AI Tools Evaluation Rating (scale 1-5) by Functions and Price Models

We also researched briefly the large language models used to support these AI tools. It appeared that most of the reviewed AI tools are supported by the GPT model (v3 free or v4 fee-based), which was the first AI LLM that went to the public, while Germin has its own LLM called

Gemini Pro 1.5, and Meta AI supported by its own LLM called Llama (v3.1). When people process text-based tasks, ChatGPT may be more preferred, while people process multimedia content or long sentence requests, Gemini may be more favored (Masaklkhhi, Ong, Waisberg, & Lee, 2024). Although Meta AI (Llama) came to the public late and only one year ago, when people process images or content related to multi-media, Mata AI (Llama) can help to provide up to 100 free images per day, in addition to its compatible functions in reading, writing and searching based on its huge amount of data sources. As Gemini integrated with its applications such as Google Chrome, Gmail, and Google Drive, Meta AI also integrated Meta AI with its social media applications such as Facebook and Instagram and with the latest information. Therefore, it is hard to comment on which LLM works better than the other (Timonera, 2024) since they all have different strengths. Besides, although GPT LLM stands alone, it provides APIs that many other applications can be integrated. On our evaluation list, other than Gemini and Meta AI, all the rest are integrated with and supported by GPT LLM.

Discussions, Conclusions, and Implications

As presented earlier, our findings addressed each of our research questions. Based on the evaluation of 50 AI tools for academic research, several recommendations and implications can be drawn. Firstly, our findings reveal a wide array of generative AI tools available to academic scholars, including the most current ones. These tools vary in their core functions, complexity, and ability to assist with tasks such as literature reviews, academic writing, and citation management. While this variety offers researchers numerous options to explore and select AI tools that best suit the different stages of academic writing, researchers should prioritize a needs-based approach to tool selection, aligning specific functionalities with research requirements. Objective evaluation criteria, encompassing accuracy, reliability, validity, complexity, usability, and efficiency/scalability, should guide this selection process. While complex, multi-functional tools, exemplified by ChatPDF, Copilot, and Bing AI proved advantageous for comprehensive research tasks, single- or dual-function tools such as Bit.ai, Research Rabbit, and Elicit demonstrated efficacy in addressing specific research needs.

Secondly, the evaluation data demonstrated that many AI tools performed well in terms of reliability and validity, with acceptable accuracy across various platforms. However, only a few AI tools offered comprehensive features. We identified three primary pricing models: freemium, fee-based, and free. Of these, the freemium model appeared to be the most widely adopted. In such cases, cost considerations warrant significant attention. The study revealed that the freemium model represents the most predominant pricing structure, but researchers should recognize that fee-based models frequently offer enhanced quality and more extensive feature sets.

Finally, ethical implications surrounding AI utilitization in academic writing necessitate careful consideration. Transparency in AI deployment and adherence to evolving journal practices regarding using AI tools like ChatGPT, ChatPDF, Copilot Sidebar, Bing AI, and Gemini in publishing are paramount. These AI tools were noted for their rich functionalit, particularly ChatPDF and Copilot, which excelled in performing internet searches and responding to prompts in an interactive, conversational manner, producing valid and accurate outputs. As highlighted earlier, three tools—ChatPDF, Copilot, and Bing AI—stood out, receiving near-

perfect evaluation scores largely due to their higher complexity. However, this does not imply that the other tools were of lower quality. A range of single- or dual-function tools also distinguished themselves with high scores in accuracy, reliability, or validity, and were noted for their ease of use. Examples include Bit.ai, Research Rabbit, LitMaps, Scite AI, Elicit, PDFgear, AI Writer, Pro Dream, and Quiltbot. These tools offer excellent performance based on specific user needs. For instance, Elicit, powered by multiple models, allows users to create custom prompts and presents results in a clear, tabular format. This functionality is especially valuable for querying multiple papers simultaneously, greatly facilitating the research process.
Our findings suggest that AI tools with more complex features are generally more preferred for assisting with academic writing, although some single- or dual-function tools also proved helpful for specific academic tasks. Cost is another factor to consider when selecting an AI tool, with the freemium pricing model proving to be more popular than fee-based or free models. When selecting AI tools, users should carefully consider their primary research needs and select tools according to their needs, using evaluation scores as we presented here or other published best practice as a guide.

As shown in our literature review earlier, only a limited number of studies have evaluated the effectiveness of AI tools in a comprehensive scope, and very little studies evaluated AI tools' quality from user-end perspective objectively. They either provided overview of other researchers' results in AI (e.g., Wagner, Lukyanenko, and Paré, 2021; Abdul, Mathew, Ahmad Saad, and Alqahtani, 2021; Khalifa and Albadawy, 2024) or focused on users' perspective about their subjective opinions (e.g., Marzuki, Widiati, Rusdin, Darwin, and Indrawati, 2023; Dwivedi, Kshetri, Slade, Jeyaraj, and Kar, 2023). Thus, this study has made a significant contribution to addressing this gap by providing not just comprehensive review but also an objective evaluation. Additionally, with the release of ChatGPT in November 2022, many studies published prior to that date have become outdated. This paper updates existing research by including newer AI tools not evaluated in those studies prior ChatGPT release, such as Wagner, Lukyanenko, and Paré (2021) study.

Furthermore, this paper enhances our understanding of how most current AI tools can improve the efficiency of the academic research process via a very structured evaluation. Particularly, based on what we found as presented earlier, we are very confident to claim that this study has enriched the studies about how newly published AI tools are changing the research academics landscape, how these various AI tools can potentially assist academic work more effectively and efficiently, and how we may evaluate their effectiveness carefully and constructively when we adopt these AI tools into our daily research work.

In addition, as explained earlier, each AI tool was rigorously tested using carefully designed metrics or guidelines, with results meticulously documented for further analysis. The evaluation methods included both quantitative measures (e.g., a 1-5 Likert scale for performance benchmarks and error rates) and qualitative feedback from users (e.g., ease of use and satisfaction levels). Also, the data was collected cross-tester for each evaluation task. This comprehensive approach provided a well-rounded assessment of each tool's strengths and weaknesses, offering a robust basis for comparison and evaluation. Such rigorous methodology supported the claim that the findings are valuable.

Recommendations

It is important to clarify that our assessment of reliability, validity, and accuracy focused primarily on the performance of specific tasks assigned to the AI tool—such as summarizing articles, providing in-depth analysis, or checking grammar—rather than verifying whether the papers identified by the AI tool actually existed, as highlighted and studied by Haman and Školník (2023). Their concerns about low accuracy in this regard merit further investigation, which could be the subject of a separate evaluation study. We hope this clarification prevents any misunderstanding regarding the scope of our accuracy-related evaluations.

Although we made efforts to mitigate bias, including averaging cross-tester scores, the subjective nature in deciding evaluation ratings could still pose a risk of bias. Therefore, we recommend future research to re-evaluate the AI tools as we assessed to either confirm or expand upon our findings. Future research should focus on validating the research findings through further investigation, refining evaluation methodologies, and critically examining the broader ethical ramifications of AI integration within academia. We hope that our evaluation methodologies will become more mature, after further verification and investigation, when it guides end-users and fellow researchers in selecting, assessing, and using appropriate AI tools to support their academic research.

Lastly, we want to emphasize the importance of addressing ethical concerns when using AI tools, although this is not the major focus of this study. We completely agreed that it is crucial to remain transparent about the extent of AI use when writing academic papers. Plagiarism through using AI assistance in writing has been a common concern in both academic research and educational practice. We encourage researchers to use AI-research tools mindfully and stay informed about the policies many academic journals have already implemented regarding AI use in scientific writing and publishing (Ciaccio, 2023).This is an important area that deserves further research.

Disclaimer

The authors of this paper declare that they have no commercial relationships or financial sponsorships with any of the AI tools discussed in this study. All evaluations and analyses presented are conducted independently, and the findings are based solely on the empirical data and research methodologies employed. The authors have no vested interest in promoting or endorsing any specific AI tool or company.

While writing this paper on AI tools for scholars, a few AI tools, including ChatGPT, were utilized to assist with the wording choices or grammatical checking occasionally. However, the primary arguments and findings are entirely grounded in the authors' research, conducted with the assistance of a group of student research team. All conclusions have been rigorously validated by the authors to ensure accuracy and integrity.

References

At least 10% of research may already be co-authored by AI (2024, June 26). *The Economist*. Retrieved from: https://www.economist.com/science-and-technology/2024/06/26/at-least-10-of-research-may-already-be-co-authored-by-ai.

Abdul Razack, H. I., Mathew, S. T., Ahmad Saad, F. F., & Alqahtani, S. A. (2021). Artificial intelligence-assisted tools for redefining the communication landscape of the scholarly world. *Science Editing, 8*(2), 134-144. https://doi.org/10.6087/kcse.244

Aydın, Ö. (2023). Google Bard Generated Literature Review: Metaverse. *Journal of AI,* 7(1), 1-14. https://doi.org/10.61969/jai.1311271

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age. *Business and Professional Communication Quarterly*, *86*(3), 257-295. https://doi.org/10.1177/23294906231176517

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics, 2*(3). https://doi.org/10.1016/j.rmal.2023.100068

Ciaccio, E. J. (2023). Use of artificial intelligence in scientific paper writing. *Informatics in Medicine Unlocked, 41*, 101253. https://doi.org/10.1016/j.imu.2023.101253

Cribben, Ivor and Zeinali, Yasser, The Benefts and Limitations of ChatGPT in Business Education and Research: A Focus on Management Science, Operations Management and Data Analytics (March 29, 2023). Available at SSRN: https://ssrn.com/abstract=4404276 or http://dx.doi.org/10.2139/ssrn.4404276

Dashti, M., Londono, J., Ghasemi, S., & Moghaddasi, N. (2023). How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing? *The Journal of Prosthetic Dentistry*. https://doi.org/10.1016/j.prosdent.2023.05.023

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Guler, N., Kirshner, S., and Vidgen, R. A literature review of artificial intelligence research in business and management using machine learning and ChatGPT (August 14, 2023). UNSW Business School Research Paper Forthcoming, Available at SSRN: https://ssrn.com/abstract=4540834 or http://dx.doi.org/10.2139/ssrn.4540834

Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. *Accountability in Research - Ethics, Integrity and Policy* 1–9. https://doi.org/10.1080/08989621.2023.2168535

Haman, M., & Školník, M. (2023). Using ChatGPT to conduct a literature review. *Accountability in Research*, 1–3. https://doi.org/10.1080/08989621.2023.2185514

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30-50. DOI:10.1007/s11747-020-00749-9

Johnson, C. D., Bauer, B. C., & Niederman, F. (2021). The Automation of Management and Business Science. *Academy of Management Perspectives*, *35*(2), 292–309. https://doi.org/10.5465/amp.2017.0159

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update, 5*(1), 100145. https://doi.org/10.1016/j.cmpbup.2024.100145

Kurniati, E. Y., & Fithriani, R. (2022). Post-graduate students' perceptions of Quillbot utilization in English academic writing class. *Journal of English Language Teaching and Linguistics, 7*(3), 437-451. https://doi.org/10.21462/jeltl.v7i3.852

Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2). https://doi.org/10.1080/2331186X.2023.2236469

Mandai, K., Tan, M. J. H., Padhi, S., & Pang, K. T. (2024). A cross-era discourse on ChatGPT's influence in higher education through the lens of John Dewey and Benjamin Bloom. *SSRN*. https://ssrn.com/abstract=4755622 or http://dx.doi.org/10.2139/ssrn.4755622

Masaklkhhi, M., Ong, J., Waisberg, E., & Lee, A.G. (2024). Google DeepMind's Gemini AI versus ChatGPT: a comparative analysis in ophthalmology. 38 (1412–1417), retrieved from https://www.nature.com/articles/s41433-024-02958-w

Mishra, H. R. (2024). Exploring the impact of artificial intelligence tools in engineering pedagogy: A qualitative survey of academic experiences. *International Journal for Research in Applied Science and Engineering Technology, 12*(1), 60-66. https://doi.org/10.22214/ijraset.2023.57864

Nguyen-Trung, K., Saeri, A. K., & Kaufman, S. (2023, April 20). Applying ChatGPT and AI-powered tools to accelerate evidence reviews. OSF Preprints, 20 April, 2023. https://doi.org/10.31219/osf.io/pcrqf

Ray, P. P. (2024). Using ChatGPT for early career research scholarship. *The Clinical Teacher*, 21(3), 1–1. https://doi.org/10.1111/tct.13694

Salvagno, M., Taccone, F.S. & Gerli, A.G. Can artificial intelligence help for scientific writing?. *Crit Care* **27**, 75 (2023). https://doi.org/10.1186/s13054-023-04380-2

Somers, M. (2023, October 19). How generative AI can boost highly skilled workers' productivity. *MIT Sloan Management Review*. Retrieved from https://mitsloan.mit.edu/ideas-made-to-matter/how-generative-ai-can-boost-highly-skilled-workers-productivity

Steele, J. L. (2023). To GPT or not GPT? Empowering our students to learn with AI. *Computers and Education: Artificial Intelligence, 5*, 100160. https://doi.org/10.1016/j.caeai.2023.100160

Timonera, K. (2024). Meta AI Chatbot Quick Facts, retrieved from eWeek at: https://www.eweek.com/artificial-intelligence/meta-ai-chatbot-review/#:~:text=One%20feature%20that%20distinguishes%20the,15%20with%20Copilot's%20free%20plan

Wagner, G., Lukyanenko, R., & Paré, G. (2021). Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology*, *37*(2), 209-226. https://doi.org/10.1177/02683962211048201

Acknowledgement

We extend our gratitude to the following student researchers for their invaluable assistance in collecting evaluation data for this research project:

Ajay Singh
Derrick Zihang Hong
Harika Kolipaka
Khaled Zaza
Mohammad Zaza
Sean Nian
Cody Ourique
Noel Hjalmarson

Warner Woes: A Case Study of Warner Bros.'s Merger & Acquisition Activity

Cory Angert, Northern Arizona University

Abstract

Mergers and acquisitions (M&As) can help firms grow and compete more effectively in the marketplace by affording businesses the opportunity to combine their operations with those of other companies. While numerous potential benefits may be realized by engaging in such a transaction, both mergers and acquisitions bear risks as well, with research's confirming a high failure rate for most M&As. Over the course of its more than century-spanning existence, media company Warner Bros. has experienced the peaks and valleys of M&A activity. The company known best for its venerable film studio, memorable characters, and innovative contributions to film, television, videogames, and other fields had survived many momentous shifts; in 2024, however, Warner Bros., having recently merged with Discovery to form Warner Bros. Discovery (WBD), once again faced a major crossroads. This study recounts the history of Warner Bros., primarily focusing on the most recent three decades, to elucidate the events that led to the company's position in Fall 2024 and then poses the question, among others, of what strategic decision the reader believes would constitute most prudent for WBD.

Keywords: acquisition, antitrust, case study, merger, strategy

Mergers and Aquisitions

Mergers and acquisitions (M&As) are transactions that result in two or more organizations' combining into one. According to Junni and Teerikangas (2019), mergers can be defined as transactions wherein "a new company is formed in which the merging parties share broadly equal ownership," while acquisitions constitute transactions in which "the acquirer purchases the majority of the shares (over 50%) of another company (the 'target') or parts of it (e.g., a business unit or a division)" (p. 1). While a high percentage of mergers and acquisitions tend to ultimately come to be viewed as failures, research suggests that certain key constructs may improve the likelihood of success (Cartwright and Schoenberg, 2006). For instance, strategic fit, or the realization of synergies such as resource sharing across divisions, could potentially result in performance gains (Cartwright and Schoenberg, 2006). Additionally, the organizations' cultural compatibility and the processes employed to bring two companies together could also influence a merger's or an acquisition's chances of success (Cartwright and Schoenberg, 2006).

Successful mergers and acquisitions can help firms gain market power, realize efficiencies, enhance innovation, achieve economies of scope, and secure other advantages (Haleblian, Devers, McNamara, Carpenter, & Davison, 2009). By the same token, however, mergers and acquisitions may also result in negative consequences. For instance, such transactions may lead to increased executive and employee turnover, economic harm for certain customers, and unrealized anticipated synergies (Haleblian et al., 2009; King, Dalton, Daily, & Covin, 2003). The following case study traces the history of Warner Bros. through multiple M&A transitions before then challenging the reader to reflect on the firm's saga and place oneself in the shoes of the corporation's (at this point, Warner Bros. Discovery) top leadership to determine the optimal path forward during a period when the organization faced a pivotal strategic juncture.

Case

(AOL) Time Warner Era

In 1923, Albert, Harry, Jack, and Sam Warner founded Warner Bros. Pictures, Inc., one of the pioneer movie studios (Hall, 2024). That same year, Briton Hadden and Henry Luce launched *Time* magazine (Smithsonian National Postal Museum, 2024). Both WB and Time, Inc. underwent multiple mergers, acquisitions, and divestments before the two firms eventually merged in 1990 to form Time Warner (Companies History, 2014). Following the merger, Time Warner continued to grow, acquiring the Turner Broadcasting System and other properties, launching the Cable News Network (CNN), and pursuing multiple other innovative initiatives (Companies History, 2014). At the turn of the century, however, the corporation experienced a major setback. At the height of the dotcom bubble circa 2000, a period during which rampant stock market speculation resulted in gross overvaluation of many internet-based companies (CFI Team, 2016), America Online (AOL) set its sights on merging with Time Warner (Lumb, 2015). Unfortunately, as soon as the two organizations combined to form AOL Time Warner in 2001, serious problems began to become abundantly apparent, leading Jeff Bewkes, the Time Warner executive who would go on to later become Time Warner CEO, to dub the transaction "the biggest mistake in corporate history" (Companies History, 2014, AOL Time Warner merger section). After a series of damaging circumstances and setbacks, including clashing corporate

cultures and broadband internet's usurping AOL's primary offering of rapidly-obsolescing dial-up internet (Lumb, 2015), in 2009, the company's leadership was forced to spin off AOL as an independent enterprise (Companies History, 2014).

WarnerMedia Era

The most recent stage of the entity formerly known as Time Warner begins with telecommunications giant AT&T's $85+ billion acquisition of the company and subsequent rebranding of Time Warner as WarnerMedia (Hall, 2024). Despite the transaction's being delayed as a result of the United States Department of Justice's suing on the basis of antitrust concerns (Hall, 2024), the deal ultimately secured approval, and WarnerMedia became a subsidiary of AT&T in June 2018 (Maas, 2022a). Four months later, the organization shared preliminary plans for a streaming video service that would come to be known as HBO Max and, subsequently, simply Max (Maas, 2022a). Given WarnerMedia's vast library of valuable and highly sought-after intellectual properties (Mazumdar 2016), this new streaming service was to become a cornerstone of AT&T's corporate strategy. AT&T would hopefully now be able to compete with Netflix, the juggernaut at the time (Spangler, 2018); realize lucrative and strategically advantageous synergies across its divisions; and pave a pathway toward what most saw, in the wake of traditional cable and satellite packages' beginning to give way to subscription video on demand (SVOD) services and other streaming alternatives, as the future of entertainment.

The firm began to invest heavily in this promising new venture, inking megadeals with the likes of J.J. Abrams (a deal valued at $500 million dollars) and others (Maas, 2022a) and making bold decisions such as pricing HBO Max at $14.99 USD. This could be viewed as an aggressive pricing scheme, because the Home Box Office (HBO) premium cable channel package alone was priced at $14.99, and this new HBO Max service included all HBO content plus a vast array of additional programming. In fact, those who already subscribed to standard HBO through their cable or satellite providers were automatically granted instant unfettered access to the HBO Max app at no additional charge, a tactic that allowed the streaming provider to establish a respectable foothold in the streaming space practically overnight. If AT&T had hopes of competing with the likes of incumbent rivals such as Netflix, leadership knew that leveraging an existing subscriber base and expanding upon this core would be essential.

HBO Max launch. When HBO Max launched in May 2020, it did so in an environment upended by the global COVID-19 pandemic. While this meant that many people were secluded in their homes, potentially presenting an audience more receptive to a service that promised hours of in-home entertainment, it also compromised AT&T's rollout. The American sitcom *Friends* had solidified its spot as one of the most-streamed programs on Netflix; so, when the license expired, WarnerMedia, which had produced the series under then-Time Warner's Warner Bros. Television division and thus owned the distribution rights, opted not to renew the contract with Netflix (Frank, 2019). Instead, *Friends* was positioned as one of the crown jewels of the HBO Max launch that the company planned to promote with a highly-anticipated *Friends* reunion special that would be available exclusively on HBO Max. Unfortunately, COVID protocols made it impossible to film the special, so AT&T was forced to launch HBO Max without this vital piece of the firm's content offering strategy (Andreeva, 2020). In addition to this impediment,

the company also had to contend with the challenges of employees' needing to work from home; app availability issues largely stemming from negotiation disputes with Roku and Amazon, two of the most prominent streaming hardware suppliers; market confusion resulting from the availability of multiple legacy apps (HBO Go and HBO Now); and numerous technical glitches and interface quirks (Marshall, 2020; Szalai, 2020).

Despite multiple setbacks, AT&T boasted about its securing a user install base of 4.1 million within the first month (Szalai, 2020). While many of these customers were simply existing HBO subscribers who were now granted access to the HBO Max app and its catalogue, the company viewed the launch numbers as a win, going so far as to call the initial rollout a "flawless launch," a laudatory statement that many would refute (Katz, 2020a; Katz, 2020b; Szalai, 2020). In light of Disney's announcing in August 2020 that its recently-launched Disney+ streaming service had acquired more than 60.5 million subscribers in just its first eight months of existence, the HBO Max metrics looked particularly paltry (Hayes and Hipes, 2020). This may be the reason, at least in part, that mere days later "recently appointed WarnerMedia CEO Jason Kilar dismissed executives Bob Greenblatt and Kevin Reilly in a shocking hierarchal restructuring of the global conglomerate" (Katz, 2020b, para. 1). Despite outwardly painting a rosy picture of the major strategic initiative's success, signs pointed to HBO Max's not meeting AT&T's internal projections.

It did not help that, with the COVID-19 pandemic's keeping moviegoers away from theaters, new film debuts all but ceased. This meant that not only were WarnerMedia's Warner Bros. movie studio and its subsidiaries unable to generate revenue from theatrical releases but also that these films would not then make their way to HBO Max following the ends of their theatrical runs. To combat this serious predicament, WarnerMedia boldly announced Project Popcorn – a commitment to launch WB's entire 2021 slate of films (plus *Wonder Woman 1984* on Christmas Day 2020), all of which were originally intended exclusively for theaters and only later destined for streaming, simultaneously in both cinemas and on HBO Max (Fink, 2023). While this unprecedented move did somewhat help to boost HBO Max's subscriber count, it also ruffled many feathers in Hollywood, especially since most of the talent responsible for the motion pictures included in Project Popcorn were not given any advance warning of the major strategic shift (Fink, 2023). Some creators even filed lawsuits against WarnerMedia and parent company AT&T as a result of the hostilities engendered by Project Popcorn; prominent filmmakers, such as director Christopher Nolan, parted ways with WB despite the collaborators' having had enjoyed longstanding and highly fruitful relationships, while other creatives seemingly lost trust in WB and chose to distance themselves from the once-venerable filmmaker-centric studio (Fink, 2023). In an effort to appease resentments and address legitimate claims of breached contracts, the corporation reportedly paid out approximately $200 million to cover backend deals and other alleged losses of revenue resulting from suppressed box office earnings, but the damage had already been done, and the trust that WB had spent years cultivating within the Hollywood community had already been broken (Fink, 2023).

Many blame Project Popcorn as the impetus behind AT&T's decision to divest WarnerMedia only three years after acquiring it (Fink, 2023). The money lost on WarnerMedia's cannibalizing its own 2021 movie ticket sales, combined with other difficulties such as falling short in successfully leveraging WarnerMedia's assets across AT&T's broader portfolio, prompted

AT&T to engage in merger negotiations with Discovery Inc. (Fink, 2023). AT&T spun off WarnerMedia so that this standalone entity could then merge with Discovery (Maas, 2022a). In April 2022, the deal was finalized, and the new company was named Warner Bros. Discovery, Inc., with the chief executive officer of Discovery appointed as CEO of WBD (Maas, 2022b).

Warner Bros. Discovery Era

To some, the combination of WarnerMedia and Discovery did not seem a natural fit, given Warner's overall emphasis on scripted, often costly, television and film projects versus Discovery's focus on cheap-to-produce unscripted series (Masters, 2021). On the heels of the dissatisfaction that many WarnerMedia employees and partners felt under AT&T's regime, however, most welcomed the change of leadership ushered in by Discovery CEO David Zaslav, who took control of the merged company (Masters, 2021). Those familiar with Zaslav largely painted him as a shrewd and ruthless businessman, ascribing him a reputation as a vicious bargainer and fearsome boss "known to yell and curse at even high-level underlings" (Masters, 2021, para. 9). Although such characterizations painted a prickly picture of the executive, Zaslav did boast formidable bona fides. Originally a lawyer by trade, Zaslav came to the realization that practicing law did not suit him, and he joined the National Broadcasting Company (NBC) in 1989 (Masters, 2021). There, he played a leading role in the launch of business news channel CNBC and also helped to establish political commentary and news network MSNBC (WBD, 2024a). In his time with NBC, subsequently renamed NBCUniversal following a merger, Zaslav oversaw numerous networks and media properties as he rose through the ranks, affording him the experience necessary to take over as CEO of Discovery in 2006 (WBD, 2024).

During Zaslav's Discovery tenure, several of the company's educational cable channels were transformed into homes for reality series that often drew criticism as being trivial, at times even exploitative, "junk food" TV (e.g. Kelley, 2023; Roth, 2023). Although shows such as Dr. *Pimple Popper*, *90 Day Fiancé*, and *Naked and Afraid* drew the ire of critics, the formula proved successful with viewers, as Zaslav recognized that a sizable market existed for what he referred to as "lean back, comfort viewing" (Francisco, 2022, What the Future Holds section; Hughes, 2022). This type of content was earmarked as an integral piece of WBD's strategy moving forward, particularly with regard to the company's streaming business initiatives. In the domain of subscription streaming, churn – the rate at which subscribers to a service come and go – serves as a prominent indicator of a service's health. If customers sign up to watch a prestigious, buzzworthy show for the two months or so that the series airs, for example, they may then cancel their memberships once the program ends if there is nothing to retain their interest. The strategy of WBD's subscription video on demand (SVOD) offering, then, would be to combine WarnerMedia's primarily scripted series – particularly critically-acclaimed HBO shows such as *Game of Thrones* and *Succession* – with Discovery's plentiful library of low-cost reality shows such as *House Hunters* and *Chopped* into a unified content package in an effort to reduce churn (Spangler, 2022). Notable award-winning series such as HBO's *The White Lotus* would lure viewers and then, while audiences waited for the next marquee series to air, they would be compelled to renew their subscriptions because they (or someone else in their households) would be hooked on *Diners, Drive-Ins, and Dives* or one of Discovery's other "sticky" evergreen shows featuring a seemingly endless supply of bingeable episodes. Essentially, WBD's bid was that the

strengths of the HBO Max and Discovery+ streaming services would complement each other by filling in the gaps in the other's content library.

One potential challenge that the strategy presented is that David Zaslav's history of running a company known for pumping out an endless supply of content placed a premium on quantity over quality and cost-consciousness over artistic integrity. As *The Hollywood Reporter* put it, "[Zaslav had] boasted that Discovery's average cost of content was $400,000 to $450,000 an hour while others were paying $5 million or more (often much more). Scripted programming is 'the red carpet, it's the sexy actors and actresses, it's the opening and it's all the glare and all the glamour,' he said then. 'That's not us.' But it is now" (Masters, 2021, para. 14). Once Zaslav took the reins of Warner Bros. Discovery, WarnerMedia channels did, indeed, cut back on the hours of scripted programming produced, including the complete cessation of scripted programming development at flagship Turner networks Turner Broadcasting System (TBS) and Turner Network Television (TNT), but this was only the first of many controversies to come (Maas and Otterson, 2022). David Zaslav would soon come to be called, in an article that he allegedly forced the publisher to retract because it cast him in such an unflattering light, "perhaps the most hated man in Hollywood" (Valdez, 2023, para. 5).

Controversial decisions. Early on, Zaslav courted controversy by immediately shuttering the just-launched CNN+ news streaming service (Casey, 2022) and then removing a swath of shows from the HBO Max streaming service (Radulovic, 2022). Many of the deleted series allegedly had modest viewership, but several were beloved by fans who felt that the extra step of scrubbing the existence of these shows from even the company's social media channels was an especially cruel measure that insulted both the shows' creators and their fanbases (Lang, 2022). Needless to say, this striking and unprecedented purge alienated both HBO Max subscribers and the general public, but the decision to scrap two nearly-completed films sent shockwaves through the entertainment industry. *Batgirl* and *Scoob! Holiday Haunt*, two movies originally commissioned as HBO Max-exclusive releases, were unceremoniously terminated in favor of WBD's receiving tax write-downs (Gonzalez, 2022). Despite filming on the approximately $90 million *Batgirl*'s having already been completed and almost all work on the roughly $40 million *Scoob! Holiday Haunt*'s having ended (the finishing touches were actually carried out after the project was cancelled even though it would never be released), Zaslav's decree that no films skip theaters and go directly to streaming meant that these releases targeted for HBO Max no longer fit into the firm's strategy (Gonzalez, 2022; Muñoz, 2023). While WBD leadership claimed that these cancellations were one-time measures taken as a result of tax reporting opportunities offered by the unique circumstances surrounding the merger, their reassurances would ring hollow when the firm later repeated these ethically dubious tactics (Kit and Couch, 2023).

The controversies did not end there. Reports surfaced that Zaslav intended to implement a programming shift across the entire company toward appealing more to middle America, eschewing previous corporate efforts to elevate minority voices by embracing diverse creators and underrepresented stories (Manno, 2022). These reports were refuted by the company, but the elimination of diversity, equity, and inclusion initiatives – and WBD's subsequent backtracking of this controversial action in response to intense backlash – and cancellation of popular LGBTQ+ and inclusive series have seemingly borne out the truth of WBD's courting broad

audiences at the expense of serving traditionally-overlooked cultural niches (Bundel, 2024; TheGrio Staff, 2022).

Zaslav's appointing a new head of cable news network CNN with the modus operandi of appealing more to conservative audiences by all accounts failed spectacularly (Alberta, 2023; Coster, 2023). The disruptive CEO then attempted to gut Turner Classic Movies (TCM), a channel intended to preserve and make accessible cinematic history, until irate filmmakers, whose support a major Hollywood studio such as Warner Bros. needs for survival, forced Zaslav to walk back this measure (Masters, 2023). Meanwhile, the company's gaming division, Warner Bros. Games, came under fire for allegedly predatory downloadable content practices (Makar, 2023) and problematic game launches (Corden, 2024; Gosnell, 2022; McWhertor, 2023). The merging of HBO Max with Discovery+ also did not go as smoothly as the company would have liked, with technical glitches and outcry as even more programming was delisted from the platform (Foreman and Chapman, 2023; Spangler, 2023). The ridicule that followed the announcement that the combined streaming service would simply be known by the exceedingly generic moniker Max may have also taken some luster out of the launch (Bissada, 2023; Tinoco, 2023). WBD ended 2023 with news that the upcoming anticipated *Looney Tunes* film *Coyote vs. ACME* would, like *Batgirl*, *Scoob! Holiday Haunt*, and possibly other projects not known to the public, be discarded so that the company could realize another tax write-off (Taylor, 2024). The backlash was swift and forceful. As a result, WBD vowed to shop the film to other studios (Taylor, 2024). Unfortunately, a few months later, sources revealed that this promise was likely only made as lip service, a ruse orchestrated to assuage the public – and perhaps the United States Congress – so that WBD could avoid further scrutiny while quietly adhering to its original plan of destroying the film (Taylor, 2024).

For many observers, the actions of David Zaslav are all the more galling given that he is consistently listed as one of the most overpaid CEOs, having topped the 2022 As You Sow list of overcompensated corporate executives and leading former Warner partner *Time* to write "At the top of the list: Warner Bros. Discovery's David Zaslav, who received $246 million in 2022 even though the company's stock fell 60% in the same year and roughly 40% of shares voted against his pay package" (Popli, 2023, para. 4). Still, WBD has enjoyed some major wins during Zaslav's tenure. In the summer of 2023, Warner Bros. Pictures' *Barbie* became a cultural touchstone when its release coincided with Universal Pictures' *Oppenheimer* to create the major grassroots event "Barbenheimer," which saw audiences' flocking to theaters in droves for the biggest double feature since before the COVID-19 pandemic (Thompson, E, 2023). (Ironically, *Oppenheimer* could have also been a WB film had Project Popcorn not alienated its director Christopher Nolan.) *Barbie* went on to win multiple awards and become the year's highest-grossing film (Boxoffice Staff, 2023). Additionally, WBD ended 2023 with the best-selling game of the year, Harry Potter title *Hogwarts Legacy*, unseating the *Call of Duty* franchise from its long-held top spot, and *Hogwarts Legacy* was joined by WB Games's *Mortal Kombat 1* in the number eight slot (Makuch, 2024). Further, on the television front, WBD trumpeted its success by proclaiming "Warner Bros. Discovery Networks Premiered 20 of the 25 Highest-Rated Freshman Series in 2023—The Most of Any Cable Portfolio" (WBD, 2024b) and "Max Receives 31 Primetime Emmy Awards, The Most of Any Network or Platform, Across 11 HBO Original Series" (WBD, 2024c). Finally, by the end of 2023, Max had become home to live news and sports broadcasts, and the service had at last become modestly profitable (Bouma, 2023; Eddy,

2023; Johnson, 2023). Other industry players took note of WBD's successes, and certain parties began to express interest in the company.

WBD's Merger and Acquisition Prospects

On December 19, 2023, David Zaslav met with Paramount Global CEO Bob Bakish (Fischer, 2023). The two executives discussed a potential merger, although later reporting suggested that the meeting's true goal may have been to force NBCUniversal, a subsidiary of Comcast and distributor of the aforementioned *Oppenheimer*, to engage in negotiations by creating a sense of urgency (Fischer, 2023; Sherman, 2023). At the beginning of 2024, NBCU parent company Comcast was rumored to be exploring entry into the lucrative videogame industry (Strickland, 2023) after years of periphery involvement in the form of efforts such as eSports sponsorships and a gaming-centric network called G4 (Comcast, 2023; Hayes, 2022). As the only major media company that holds both a top movie studio and a leading videogame studio within its portfolio, WBD would appear a ripe target for M&A activity should Comcast, indeed, seek significant entry into gaming. In fact, as early as mid-2022, when WBD first formed, rumors were swirling that the long game was for WBD to ruthlessly cut costs and pay down debt ahead of a merger or acquisition with Comcast [once the mandatory waiting period between M&As for companies seeking to benefit from a Reverse Morris Trust (RMT) expires two years after a merger or acquisition has finalized (in this case, mid-2024)] (Li, 2022; Siegel, 2024; Ward, 2021).

Zaslav insisted that WBD was not for sale and would instead focus on leveraging its franchises – the newly-rebooted DC Comics universe, the *Game of Thrones* series, and the world of Harry Potter, chief among them – and videogame development and publishing capabilities to achieve success on its own (Maas, 2024; Siegel, 2024). Still, the possibility of the corporation's potentially engaging in M&A with another firm presented an intriguing prospect that merits examination, discussion, and speculation. As of 2024, WBD stood poised to combine forces in some manner with another media giant; below is a table highlighting WBD's attributes and those of some possible partners.

In the United States, companies typically face scrutiny when attempting to engage in major merger or acquisition activity. The Sherman Antitrust Act of 1890 rendered illegal certain types of anti-competitive cartel-like actions and monopolization, while 1914's passages of the Clayton Act and the Federal Trade Commission Act helped bolster antitrust protections (Kovacic and Shapiro, 2000; Shapiro, 2018). While courts' and government officials' interpretations of what constitutes illegal anti-competitive moves has fluctuated over time, the intention of fostering competition by combating collusion ostensibly remains the objective of U.S. antitrust policy (Kovacic and Shapiro, 2000). Since the 1970s, a confluence of factors including political pressure, regulatory agency budget cuts, and changes in the Supreme Court's composition has resulted in the pendulum's largely swinging toward loosened enforcement of antitrust principles (Lancieri, Posner, & Zingales, 2023). While some have urged that the U.S. government employ stronger antitrust regulation and enforcement (e.g. Shapiro, 2018), multibillion-dollar acquisitions and so-called "megamergers" remain relatively commonplace in the current business environment (Khanna, 2022).

When AT&T first expressed its intention to purchase Time Warner, the news raised some red flags at the U.S. Department of Justice (DOJ) (Maas, 2022a; U.S. Department of Justice, Office of Public Affairs, 2017). Although then-president Donald Trump's personal views towards Time Warner-owned news network CNN perhaps played a role in the DOJ's legal challenge (Gold, 2019), the DOJ argued that the transaction, which ranked among the highest-valued combinations of two companies in the history of the United States, would materially harm both consumers and competitors (U.S. Department of Justice, Office of Public Affairs, 2017). Nevertheless, AT&T – no stranger to accusations of antitrust violations, notoriously being forced in 1982 to break up into multiple companies as a result of a landmark DOJ win (MacAvoy and Robinson, 1983) – prevailed in court, gaining unconditional approval to proceed with its planned merging of the two firms (Salinas, 2018). AT&T's decision to divest Time Warner (now called WarnerMedia) and spin it off into a merger with Discovery, a mere three years later, again raised some eyebrows (Feiner, 2021), but the controversial actions initiated by the CEO of the newly-merged company, Warner Bros. Discovery, attracted even more scrutiny. (See Table 1).

Table 1. Comparison of WBD and Potential Partners

	Warner Bros. Discovery	Comcast (NBCU)	Paramount	Disney	Sony	Netflix	Amazon	Apple
SVOD service(s)	Max & B/R Sports	Peacock	Paramount+	Disney+, Hulu, & ESPN+	x	Netflix	Prime Video	Apple TV+
Other Streaming Service(s)	TBA FAST service	Xumo Play FAST service	Pluto TV FAST/AVOD service	x	x	x	Freevee AVOD service	x
Film Production & Distribution	✓	✓	✓	✓	✓	✓	✓	✓
Television Production	✓	✓	✓	✓	✓	✓	✓	✓
OTA U.S. Network Distribution	x	✓	✓	✓	x	x	x	x
Basic Cable Television Distribution	✓	✓	✓	✓	x	x	x	x
Premium Cable Television Distribution	HBO & Cinemax	x	Showtime & The Movie Channel	x	x	x	MGM+	x
Videogame Production & Distribution	✓	x	x	x	✓	✓	✓	Distribution only
Videogame Hardware Manufacturer	x	x	x	x	✓	x	x	✓
Music Production & Distribution	✓	✓	✓	✓	✓	x	x	Classical music only
Assorted Signature Intellectual Property	DC, Game of Thrones, Harry Potter, Mortal Kombat, Dune, The Matrix, Looney Tunes, Rick and Morty, Friends, The Sopranos, The Bachelor, Sesame Street	The Fast and The Furious, Jurassic Park, Shrek, Minions, Trolls, The Office, Saturday Night Live, Suits, Law & Order, Real Housewives	Mission: Impossible, Top Gun, Star Trek, South Park, SpongeBob SquarePants, Teenage Mutant Ninja Turtles, Yellowstone, The Twilight Zone, MTV	Marvel (except Spider-Man), Star Wars, Pixar, The Muppets, The Simpsons, Avatar, Alien, Predator, Planet of the Apes, Die Hard, Ice Age, Home Alone	Spider-Man, Jumanji, Men in Black, Uncharted, The Last of Us, The Karate Kid, Resident Evil, God of War, Ratchet & Clank	Stranger Things, The Witcher, Squid Game, Wednesday, Bridgerton, Orange is the New Black, Money Heist, You, Bird Box, Extraction, The Kissing Booth	James Bond, Rocky, Stargate, Legally Blonde, Barbershop, The Boys, Invincible, Reacher, Jack Ryan, Bosch, Mr. & Mrs. Smith, Jury Duty	Ted Lasso, The Morning Show, Severance, Mythic Quest, Peanuts, Fraggle Rock

SVOD = subscription video on demand
AVOD = advertising-based video on demand
FAST = free advertising-supported streaming television
OTA = over the air broadcasts
✓ = possessed by firm
x = not possessed by firm

WBD CEO David Zaslav stunned the entertainment industry in August 2022 when the firm announced the cancellations of *Batgirl* and *Scoob! Holiday Haunt* despite both films' being nearly complete (Gonzalez, 2022). In response, United States Congressman Joaquin Castro, Senator Elizabeth Warren, and other lawmakers accused WBD of "adopt[ing] potentially anticompetitive practices" (Cho, 2023, para. 2). The December 2023 announcement of *Coyote vs. ACME*'s deletion prompted Rep. Castro to compare WBD's actions to insurance fraud, writing "it's like burning down a building for the insurance money" (Thompson, J, 2023, para. 3).

The M&A activities involving Warner Bros. have thus far succeeded even in the face of heavy resistance, but some of the potential suitors for the organization, should David Zaslav decide to once again sell off WB, might finally prove a bridge too far. For example, if WBD were to target Comcast, the fact that Comcast's NBCUniversal division operates MSNBC, one of America's other leading cable news networks, could pose a challenge as a result of the consolidation's being perceived as anticompetitive. Additionally, the fact that Warner Bros. Motion Picture Group and Universal Pictures ranked as two of the top three movie studios of 2023 could similarly pose an issue (Tartaglione, 2024). It should be apparent that antitrust concerns and other potential stumbling blocks can stymie even the best-laid plans, so WBD must exercise caution and prudence as it weighs its options for the future. Navigating mergers and acquisitions can prove a most daunting task; devising astute solutions requires a keen strategic mind.

SUGGESTED QUESTIONS FOR DISCUSSION AND ANALYSIS

1. Since Warner Bros.'s founding, what do you believe to be the company's most noteworthy M&A success? What do you consider the firm's most damaging failure?
2. Which are the most pivotal decisions that brought Warner Bros. to where it found itself in 2024?
3. How does Max, WBD's main streaming platform, compare to other streaming video services? What are Max's relative strengths and weaknesses, and how would you improve Max to make it even more competitive in the marketplace?
4. WBD CEO David Zaslav has expressed a desire to lean heavily on the company's treasure trove of intellectual property (IP) in order to build lucrative franchises. Which IP do you see as being most valuable, and how would you attempt to leverage this property?
5. What is the ethicality of scrapping projects, including major motion pictures such as Batgirl and Scoob! Holiday Haunt, that are nearly complete at the time of deletion? How might key stakeholder groups be affected by the discarding of artists' creative works?
6. If you were the CEO of WBD and decided to refrain from engaging in M&A activity, what alternative strategy would you contemplate pursuing?
7. If you were the CEO of WBD and decided to pursue a merger or acquisition with one of the companies listed in Table 1, what antitrust concerns might you face?
8. If you were the CEO of WBD, what would you do – would you seek to engage in a merger or acquisition? Why or why not?
9. If you were the CEO of WBD and had to choose a company with which to pursue an M&A strategy, which company would you select? What synergies do you identify between the two firms? What challenges might the two firms face in trying to combine into one?

10. If you were the CEO of WBD and your company merged with, acquired, or was acquired by the firm that you identified in Question 9, what would be your strategy for the newly-combined entity going forward?

References

Alberta, T. (2023, June 2). *Inside the meltdown at CNN.* The Atlantic. https://www.theatlantic.com/politics/archive/2023/06/cnn-ratings-chris-licht-trump/674255/

Andreeva, N. (2020, August 7). *'Friends': Filming of HBO Max reunion special postponed again over COVID-19*. Deadline. https://deadline.com/2020/08/friends-filming-hbo-max-reunion-special-postponed-again-covid-19-delay-tbd-1203007352/

Bissada, M. (2023, April 12). *Warner Bros. Discovery's Max streamer ripped as a terrible brand strategy: 'Insanely bad decision.'* The Wrap. https://www.thewrap.com/hbo-max-new-name-warner-bros-discovery-reactions-twitter/

Bouma, L. (2023, October 16). *Warner Bros. Discovery says Max has become profitable & says the days of cheap streaming are over.* Cord Cutters News. https://cordcuttersnews.com/warner-bros-discovery-says-max-has-become-profitable-says-the-days-of-cheap-streaming-are-over/

Boxoffice Staff. (2023, December 31). *The ten highest-grossing movies at the global box office in 2023.* Boxoffice Pro. https://www.boxofficepro.com/the-ten-highest-grossing-movies-at-the-global-box-office-in-2023/

Bundel, A. (2024, January 23). *Max delivered the final blow to the horny, quirky comedy.* Primetimer. https://www.primetimer.com/features/hbo-max-our-flag-means-death-minx-the-flight-attendant-horny-quirky-comedies

Casey, H. T. (2022, April 21). *CNN Plus is dead – after less than a month.* Tom's Guide. https://www.tomsguide.com/news/cnn-plus-is-reportedly-dead-already-after-less-than-a-month

CFI Team. (2016). *Dotcom bubble.* Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/dotcom-bubble/

Cho, W. (2023, April 7). *Warner Bros. Discovery merger under fire from lawmakers asking Justice Dept. to revisit deal.* The Hollywood Reporter. https://www.hollywoodreporter.com/business/business-news/warner-bros-discovery-lawmakers-justice-dept-1235369365/

Comcast. (2023, October 9). *Xfinity recognized as a leader in esports and gaming.* Comcast. https://corporate.comcast.com/stories/xfinity-recognized-leader-esports-gaming

Companies History. (2014, August 1). *Time Warner.* CompaniesHistory.com. Retrieved June 1, 2024, from https://www.companieshistory.com/time-warner/

Corden, J. (2024, January 30). *Searches for 'Suicide Squad refund' surge 791% following the game's early access launch 'disaster.'* Windows Central. https://www.windowscentral.com/gaming/searches-for-suicide-squad-refund-surge-791-following-the-games-early-access-launch-disaster

Coster, H. (2023, May 18). *CNN getting more Republicans on-air as it seeks political diversity.* Reuters. https://www.reuters.com/world/us/cnn-getting-more-republicans-on-air-it-seeks-political-diversity-2023-05-18/

Eddy, C. (2023, October 10). *Max launches new live sports streaming tier.* Gizmodo. https://gizmodo.com/max-live-sports-streaming-bleacher-report-mlb-nba-games-1850911983

Feiner, L. (2021, May 17). *AT&T battled the DOJ to buy Time Warner, only to spin it out again three years later.* CNBC. https://www.cnbc.com/2021/05/17/att-fought-doj-for-time-warner-only-to-spin-out-three-years-later.html

Fink, R. (2023, February 26). *Day to date releasing: 5 biggest ways Project Popcorn hurt Warner Bros.* MovieWeb. https://movieweb.com/project-popcorn-warner-bros/

Fischer, S. (2023, December 20). *Scoop: Warner Bros. Discovery in talks to merge with Paramount Global.* Axios. https://www.axios.com/2023/12/20/warner-bros-paramount-merger-discovery-streaming

Foreman, A., & Chapman, W. (2023, May 17). *87 titles unceremoniously removed from HBO Max.* IndieWire. https://www.indiewire.com/gallery/removed-hbo-max-movies-shows-warner-bros-discovery-merger-list/

Francisco, E. (2022, August 4). *Warner Bros. Discovery CEO announces ten-year "reset" plan for DCEU.* Inverse. https://www.inverse.com/entertainment/warner-bros-discovery-ceo-david-zaslav-ten-year-reset-for-dc-movies

Frank, A. (2019, July 9). *Friends is leaving Netflix in 2020*. Vox. https://www.vox.com/culture/2019/7/9/20687923/friends-leaving-netflix-2020-streaming-hbo-max

Gold, H. (2019, March 4). *Report: Trump asked Gary Cohn to block AT&T-Time Warner merger.* CNN. https://www.cnn.com/2019/03/04/media/att-time-warner-trump-gary-cohn/index.html

Gonzalez, U. (2022, August 2). *'Batgirl' won't fly: Warner Bros. Discovery has no plans to release nearly finished $90 million film.* The Wrap. https://www.thewrap.com/batgirl-movie-dead-warner-bros-discovery-has-no-plans-to-release-nearly-finished-90-million-film/

Gosnell, J. (2022, October 20). *Gotham Knights is having a rough launch.* Game Rant. https://gamerant.com/gotham-knights-launch-when-bad/

Haleblian, J., Devers, C. E., McNamara, G., Carpenter, M. A., & Davison, R. B. (2009). Taking stock of what we know about mergers and acquisitions: A review and research agenda. *Journal of Management*. *35*(3), 469-502.

Hall, M. (2024, March 15). *WarnerMedia.* Britannica. Retrieved June 1, 2024, from https://www.britannica.com/topic/Time-Warner-Inc

Hayes, D. (2022, October 16). *Comcast pulls plug on G4 TV, ending comeback try for gamer-focused network.* Deadline. https://deadline.com/2022/10/comcast-pulls-plug-on-g4-tv-ending-comeback-try-video-game-network-1235145219/

Hayes, D., & Hipes, P. (2020, August 4). *Disney+ passes 60.5M subscribers, reaches 5-year streaming goal in first eight months.* Deadline. https://deadline.com/2020/08/disney-nears-5-year-streaming-goal-in-first-eight-months-with-57-5m-subscribers-1203003841/

Hughes, W. (2022, August 5). *"Genredoms," "male skew," and all the other dumb stuff from today's HBO Max/Discovery+ merger.* The A.V. Club. https://www.avclub.com/hbo-max-discovery-plus-genredom-male-skew-merger-1849375117

Johnson, T. (2023, September 27). *CNN Max launches with schedule that mirrors much of its linear lineup.* Deadline. https://deadline.com/2023/09/cnn-max-launches-with-schedule-that-mirrors-much-of-its-linear-lineup-1235557198/

Junni, P., & Teerikangas, S. (2019, April 26). *Mergers and acquisitions.* Oxford Research Encyclopedias, Business and Management. https://oxfordre.com/business/display/10.1093/acrefore/9780190224851.001.0001/acrefore-9780190224851-e-15

Katz, B. (2020a, June 4). *The early HBO Max numbers don't look great.* Observer. https://observer.com/2020/06/hbo-max-launch-interest-compared-disney-plus-quibi-apple-tv-plus/

Katz, B. (2020b, August 10). *The challenges WarnerMedia's CEO is facing after HBO Max's slow start.* Observer. https://observer.com/2020/08/hbo-max-subsribers-warnermedia-challenges-jason-kilar/

Kelley, A. (2023, April 24). *TLC has become more exploitative than educational.* Collider. https://collider.com/tlc-reality-tv-exploitation/

Khanna, S. (2022, April 8). *Amid COVID recovery, megamergers will be commonplace.* Moneycontrol. https://www.moneycontrol.com/news/opinion/amid-covid-recovery-megamergers-will-be-commonplace-8332531.html

King, D. R., Dalton, D. R., Daily, C. M., & Covin, J. G. (2003). Meta-analysis of post-acquisition performance: Indications of unidentified moderators. *Strategic Management Journal*. *25*(2), 187-200.

Kit, B., & Couch, A. (2023, November 9). *Warner Bros. shelves John Cena's 'Coyote vs. Acme' movie a year after it completed filming.* The Hollywood Reporter. https://www.hollywoodreporter.com/movies/movie-news/john-cena-coyote-vs-acme-movie-shelved-1235643235/

Kovacic, W. E., & Shapiro, C. (2000). Antitrust policy: A century of economic and legal thinking. *Journal of Economic Perspectives*. *14*(1), 43-60.

Lancieri, F., Posner, E. A., & Zingales, L. (2023). The political economy of the decline of antitrust enforcement in the United States. *National Bureau of Economic Research Working Paper Series*.

Lang, J. (2022, August 23). *'Infinity Train' creator calls out 'unprofessional, rude . . . slimy' tactics at Warner Bros. Discovery.* Cartoon Brew. https://www.cartoonbrew.com/ideas-commentary/infinity-train-owen-dennis-warner-discovery-220453.html

Li, J. (2022, September 29). *Warner Bros. Discovery shuts down acquisition rumors, CEO claims "we are not for sale."* Hypebeast. https://hypebeast.com/2022/9/warner-bros-discovery-shuts-down-acquisition-rumors-ceo-we-are-not-for-sale

Lumb, D. (2015, May 12). *A brief history of AOL.* Fast Company. https://www.fastcompany.com/3046194/a-brief-history-of-aol

Maas, J. (2022a, March 11). *AT&T's short, bumpy ride in Hollywood: Timeline of WarnerMedia's road to Discovery spinoff deal.* Variety. https://variety.com/2022/tv/news/att-warnermedia-discovery-timeline-1235198123/

Maas, J. (2022b, April 8). *Discovery closes acquisition of AT&T's WarnerMedia.* Variety. https://variety.com/2022/tv/news/discovery-warnermedia-merger-close-warner-bros-discovery-1235200983/

Maas, J. (2024, January 26). *Inside Warner Bros. games' big live services push and doubling down on DC, 'Game of Thrones' and more franchises.* Variety. https://variety.com/2024/gaming/news/warner-bros-video-games-live-services-dc-game-of-thrones-1235888944/

Maas, J., & Otterson, J. (2022, April 26). *Warner Bros. Discovery cuts scripted programming development at TBS, TNT.* Variety. https://variety.com/2022/tv/news/tnt-tbs-scripted-programming-development-scrapped-warner-bros-discovery-1235241348/

MacAvoy, P. W., & Robinson, K. (1983). Winning by losing: The AT&T settlement and its impact on telecommunications. *Yale Journal on Regulation*. *1*(1), 1-2.

Makar, C. (2023, November 8). *Mortal Kombat 1 continues to illustrate that gross microtransactions work.* VG247. https://www.vg247.com/mortal-kombat-1-gross-microtransactions-work

Makuch, E. (2024, January 19). *The 20 best-selling games of 2023 in the US*. GameSpot. https://www.gamespot.com/gallery/the-20-best-selling-games-of-2023-in-the-us/2900-4951/

Manno, A. (2022, August 25). *Laid-off HBO Max execs reveal Warner Bros. Discovery is killing off diversity and courting 'Middle America.'* The Daily Beast. https://www.thedailybeast.com/laid-off-hbo-max-execs-reveal-warner-bros-discovery-is-killing-off-diversity-and-courting-middle-america

Marshall, R. (2020, June 8). *HBO Max review: A rough start, but plenty of potential.* Digital Trends. https://www.digitaltrends.com/movies/hbo-max-review-hands-on-impressions/

Masters, K. (2021, May 18). *As David Zaslav takes deal crown, how will he rule his Discovery-WarnerMedia empire?* The Hollywood Reporter. https://www.hollywoodreporter.com/business/business-news/david-zaslav-discovery-warnermedia-plans-1234955316/

Masters, K. (2023, June 28). *After filmmaker outcry over TCM cuts, Warner Bros. reverses course (sort of).* The Hollywood Reporter. https://www.hollywoodreporter.com/business/business-news/zaslav-reverses-tcm-changes-1235525256/

Mazumdar, A. (2016, October 27). *AT&T's Time Warner purchase brings a fortune in IP*. Bloomberg Law. https://news.bloomberglaw.com/esg/at-ts-time-warner-purchase-brings-a-fortune-in-ip

McWhertor, M. (2023, March 27). *MultiVersus is going offline, relaunch planned for 2024.* Polygon. https://www.polygon.com/23658469/multiversus-offline-update-relaunch-date

Muñoz, D. P. (2023, January 2). *Scoob! Holiday Haunt director explained why he finished it despite cancellation.* Game Rant. https://gamerant.com/scoob-holiday-haunt-director-michael-kurinsky-finishing-film/#:~:text=Holiday%20Haunt%2C%20a%20cancelled%20film,see%20the%20light%20of%20day

Popli, N. (2023, February 16). *CEO pay was up 21% in 2022. These are the most overpaid CEOs, according to a shareholder advocacy group.* Time. https://time.com/6256076/most-overpaid-ceos-2022/

Radulovic, P. (2022, August 19). *Infinity Train, Summer Camp Island, and other shows wiped from HBO Max.* Polygon. https://www.polygon.com/entertainment/23313193/hbo-max-infinity-train-summer-camp-island-ok-ko-watch

Roth, E. (2023, April 22). *Yes, I actually pay for Discovery Plus.* The Verge. https://www.theverge.com/23693293/discovery-plus-subscription-streaming-hbo-max

Salinas, S. (2018, June 12). *AT&T wins: Judge clears $85 billion bid for Time Warner with no conditions.* CNBC. https://www.cnbc.com/2018/06/12/att-time-warner-ruling.html

Shapiro, C. (2018). Antitrust in a time of populism. *International Journal of Industrial Organization*. *61*(November 2018), 714-748.

Sherman, A. (2023, December 20). *Warner Bros. Discovery merger talks with Paramount Global may draw out NBCUniversal.* CNBC. https://www.cnbc.com/2023/12/20/warner-bros-paramount-merger-talks-nbcuniversal.html

Siegel, T. (2024, February 21). *Warner Bros. spends big: 'Joker 2' budget hits $200 Million, Lady Gaga's $12 million payday, courting Tom Cruise's new deal and more.* Variety. https://variety.com/2024/film/news/warner-bros-spending-joker-2-budget-tom-cruise-deal-1235917640/

Smithsonian National Postal Museum. (2024). *Time Inc.* Smithsonian National Postal Museum. Retrieved June 1, 2024, from https://postalmuseum.si.edu/exhibition/america%E2%80%99s-mailing-industry-industry-segments-magazine-publishers/time-inc

Spangler, T. (2018, April 17). *Netflix stock shoots to all-time high after strong Q1 earnings*. Variety. https://variety.com/2018/digital/news/netflix-stock-all-time-high-q1-earnings-beat-1202756035/

Spangler, T. (2022, August 4). *HBO Max, Discovery+ to merge into single streaming platform starting in summer 2023.* Variety. https://variety.com/2022/digital/news/hbo-max-discovery-plus-merger-2023-1235333314/

Spangler, T. (2023, May 23). *Max not working? Some users report log-in errors, crashes as HBO Max converts to new streaming platform.* Variety. https://variety.com/2023/digital/news/max-down-not-working-login-errors-1235622581/

Strickland, D. (2023, May 31). *Comcast could enter the $200 billion video games market, analysts speculate.* TweakTown. https://www.tweaktown.com/news/91682/comcast-could-enter-the-200-billion-video-games-market-analysts-speculate/index.html

Szalai, G. (2020, July 23). *HBO Max reached 4.1 million subscribers one month after launch.* The Hollywood Reporter. https://www.hollywoodreporter.com/tv/tv-news/hbo-max-reached-41-million-subscribers-one-month-launch-1303735/

Tartaglione, N. (2024, January 11). *International box office 2023: Turnstiles danced with growth in key markets while the floor for big titles dropped & China shrugged; Global studio rankings.* Deadline. https://deadline.com/2024/01/international-box-office-2023-global-studio-rankings-market-share-1235709538/

Taylor, D. (2024, February 9). *The final days of 'Coyote vs. Acme': Offers, rejections and a Roadrunner race against time.* The Wrap. https://www.thewrap.com/coyote-vs-acme-update-offers-warner-bros/

TheGrio Staff. (2022, October 13). *Warner Bros. backtracks on diversity program cuts, keeps them alive.* TheGrio. https://thegrio.com/2022/10/13/warner-bros-backtracks-on-diversity-program-cuts-keeps-them-alive/

Thompson, E. (2023, July 17). *What Is Barbenheimer? The biggest movie event of the year explained.* Us Weekly. https://www.usmagazine.com/entertainment/news/what-is-barbenheimer-the-movie-event-of-the-year-explained/

Thompson, J. (2023, November 14). *Rep. Joaquin Castro calls Warner Bros. Discovery 'predatory' for shelving 'Coyote vs. Acme,' asks Justice Dept. to 'review this conduct.'* Variety. https://variety.com/2023/film/news/joaquin-castro-warner-bros-discovery-coyote-vs-acme-1235790855/

Tinoco, A. (2023, May 23). *Peacock weighs in on HBO Max's rebranding with R-rated name change suggestion of its own.* Deadline. https://deadline.com/2023/05/peacock-weighs-in-on-hbo-max-rebranding-1235377740/

U.S. Department of Justice, Office of Public Affairs. (2017, November 20). *Justice Department challenges AT&T/DirecTV's acquisition of Time Warner.* DOJ. https://www.justice.gov/opa/pr/justice-department-challenges-attdirectv-s-acquisition-time-warner

Valdez, J. (2023, July 6). *How GQ erased its article that was critical of Warner Bros. Discovery CEO David Zaslav.* The Los Angeles Times. https://www.latimes.com/entertainment-arts/story/2023-07-06/gq-removed-its-article-warner-bros-discovery-ceo-david-zaslav

Ward, A. (2021, May 20). *Reverse Morris Trust (RMT).* Financial Edge. https://www.fe.training/free-resources/ma/reverse-morris-trust-rmt/

WBD. (2024a). *David Zaslav.* Warner Bros. Discovery. Retrieved June 1, 2024, from https://wbd.com/leadership/david-zaslav/

WBD. (2024b, January 8). *Warner Bros. Discovery networks premiered 20 of the 25 highest-rated freshman series in 2023 – The most of any cable portfolio.* Warner Bros. Discovery. https://wbd.com/warner-bros-discovery-networks-premiered-20-of-the-25-highest-rated-freshman-series-in-2023-the-most-of-any-cable-portfolio/

WBD. (2024c, January 16). *Max receives 31 Primetime Emmy Awards, the most of any network or platform, across 11 HBO original series.* Warner Bros. Discovery. https://press.wbd.com/us/media-release/hbo-and-max-receive-31-primetime-emmy-awards-most-any-network-or-platform?language_content_entity=en

Developing a Disinformation Incident Response Playbook:

Combatting Real-Time Disruption via Deepfakes and Generative AI

Edward L. Mienie, University of North Georgia

Bryson R. Payne, University of North Georgia

Abstract

We are entering an age in which disinformation, fake news, and falsified images, videos, and audio 1) are rapidly becoming indistinguishable from authentic media, 2) can be produced in real-time, and 3) can be deployed at scale and in quantities that businesses, news agencies, and nations alike may not be able to respond to effectively. Over the past four years, deepfake videos, in which an actor's face can be replaced with a believable facsimile of a CEO's or other famous person's face, have become relatively commonplace in popular culture, and deepfakes have already been used at least at a rudimentary level in disinformation campaigns. ChatGPT, a generative large-language AI model that can produce authentic-sounding human-readable text, can generate fake news articles, emails, and blog or social media posts in real-time that seem fluent and realistic to the reader. Newer generative AI tools for creating audio, video, and photorealistic images can lend additional credibility to disinformation, misinformation, and fake news and spread them online faster than human reporters and government officials can fact-check or respond. This research examines the perfect storm of disinformation enabled by these combined technologies, provides a review of existing and emerging literature in the field, and includes a brief case study on Ukraine's response to the 2022 Zelensky deepfake video at the onset of the Russian invasion to draw out recommendations for businesses, governments, and news organizations in countering AI-enhanced disruption.

Keywords: cybersecurity, identity theft, account security, multi-factor authentication

Introduction

"Falsehoods traverse the globe while veracity is still fastening its trousers."
– ChatGPT, 2023

Artificial Intelligence (AI) is a technological innovation with the potential to disrupt most aspects of human life. The era of fake news, misinformation, disinformation, and post-truth is already impacting decision-making within the realm of organizational and national security. With generative AI capable of producing realistic text, speech, images, and audio through such technologies as ChatGPT, Stable Diffusion, Midjourney, and others, added to existing deepfakes video-altering technology, the line between reality and machine-generated misinformation has not only blurred, it has been all but erased.

A recent study by University College London found that both English and Mandarin Chinese speakers were able to correctly identify artificially generated speech only 73% of the time at 2023 levels of technology (Mai, Brai, Davies & Griffin, 2023). This means that with now-dated AI speech generation technology, 27% of users would likely believe the content of AI-generated audio was authentic. Furthermore, researchers have demonstrated that humans mistake deepfake and authentic videos as much as 66% of the time even when the two are shown side-by-side (Allen et al., 2023), and technology in this area is expected to continue to advance, making it even more difficult to discern true human speech and video from completely fabricated AI-generated media (Farah, 2023).

Newer generative AI tools are capable of creating believable audio, video, and photorealistic images that can be used to spread misinformation and disinformation faster than public relations and government officials can respond. This research outlines the converging perfect storm of disinformation enabled by these combined technologies, provides a review of existing and emerging literature related to AI-generated and AI-enhanced disinformation, and provides a brief case study on Ukraine's response to the 2022 Zelensky deepfake video at the onset of the Russian invasion. The goal of this research is to compile recommendations for businesses, governments, and even individuals in developing a disinformation incident response playbook to counter AI-enhanced disruption in real-time.

Organizational and National Security Implications

Rapidly advancing generative AI technology can reasonably have a disruptive effect on the decision-making process, both for individuals and for businesses, as well as for governments and national security. Decision makers already must contend with changing ideologies, policies, other disruptive technologies, cultural shifts, and social changes in addition to traditional adversarial threats. Artificial disruption in near real-time adds another challenging dimension to an existing and increasingly complicated, duplicitous, and counterfeit world. Deepfakes, ChatGPT, and generative artificial intelligence (a subset of machine learning, or ML) pose challenges to the established mechanisms used to inform decision-makers about world events. Ultimately, decision-makers must make the most informed and accurate decisions that will enhance our nation's national security. Today, national security advisors are being challenged by the ever-changing and increasingly

sophisticated daily advancements of AI. In the major fields of AI, quantum technologies, and advanced materials, China is the leading country in 37 of 44 technologies, producing more than five times as much high-impact research as the US, its closest competitor (ASPI, 2023).

The federal government has identified AI's possible applications for defense and intelligence and has made it a major priority. However, Tucker (2020) argues that policymakers and leaders must better understand how AI systems reach their conclusions, and before the United States Intelligence Community (IC) can use AI to its full potential, it must be hardened against attack. The Office of the Director of National Intelligence (ODNI) in 2018 launched a strategy for augmenting intelligence using machines (AIM) to foster stronger collaboration by organizing and sharing their AI efforts thereby creating a synergy across the IC. This initiative forces agencies to look outside their environments as they are so consumed within their spaces with these fast-developing AI systems. ODNI is keen to integrate the IC's many unintended information silos, and agencies are applying a more integrated approach to AI to help transform tradecraft (Shapiro, 2022). "We're really working toward a whole-agency approach toward AI," Lakshmi Raman, the CIA's chief of AI said at a recent Intelligence and National Security Alliance conference. She stated that AI technology has "relevance for data collection, analysis, digital innovation, operations, and even legal and finance areas" (Shapiro, 2022). Some of our main adversaries are also investing in AI research and development, and in our collective quest for competitive advantage, the potential to use poorly understood or untested systems could lead to serious unintended consequences that could impact the world.

The business world is wise to embrace AI technologies, as research shows that business organizations become more competitive, efficient, and innovative when doing so. However, if AI is going to serve the good of humanity, it must be applied responsibly and ethically if it is going to be the key driver for positive change as anticipated (Martinovic, Bandur, & Tusevski, 2024).

Implications for the Intelligence Community

The United States' intelligence community (IC) has been focused on AI for a long time, examining ways to leverage its power, and, by implication, give the US an advantage to set precedents that other international actors could resist, comply with, or negotiate (Moran, Burton, Christou, 2023). The Defense Advanced Research Projects Agency (DARPA) announced a $2 billion campaign to develop the next wave of AI Technologies to research more collaborative and trusting partnerships between machines and humans (DARPA, 2018). The IC recognizes that the private sector performs the lion's share of AI systems research and development and that working with the private sector poses challenges of many sorts, including conflicting interests and ideologies (Moran, et al, 2023). Moran et al (2023) posit that it is premature to talk about an intelligence revolution brought about by AI because of cultural tensions within the global AI ecosystem and local and international rules and regulations governing data collection and storage.

In 2022, The White House announced that it wanted a "Blueprint for an AI Bill of Rights," which should "protect the American people from unsafe and ineffective systems;" that they should "not face discrimination by algorithms and systems should be used and designed in an

equitable way;" that they should be "protected from abusive data practices via built-in protections and should have agency over how data about them is being used;" that they "should know that an automated system is being used and understand how and why it contributes outcomes that impact them;" and that they "should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems they encounter" (The White House, 2022). The White House in May 2023 pledged a "road map" for managing AI. The plan provides for "international cooperation to manage the impact of AI." The White House acknowledges the broad applications of AI, while at the same time recognizing that the risks that AI presents need to be managed effectively by way of regulatory intervention by governments worldwide (Milligan, 2023).

At the same time, we realize that AI technology is growing at a rate faster than regulators can respond to it. The Bipartisan Policy Center and Georgetown University's Center for Security and Emergency Technology posit that, inter alia, the US must work closely with its allies and partners while also cooperating pragmatically and selectively with its adversaries such as Russia and China; prevent the transfer of sensitive AI technologies to China through export and investment controls; and implement processes to develop and deploy defense and intelligence applications of AI systems by focusing on trustworthiness, human-machine teaming, and Department of Defense's (DOD) ethical principles for AI (Bipartisan Policy Center, 2020).

AI presents an array of opportunities for strengthening the efficiency and effectiveness of intelligence procedures and challenges to the organizational structure within the U.S. Intelligence Community (IC). AI capabilities could be 1) integrated into the collection, analysis, and management processes, 2) used to educate the IC workforce as AI will influence world political events, and 3) assist the IC in the collection, analysis, and dissemination of information on AI developments worldwide. The IC would have to dedicate time, resources, and effort to accomplish the aforementioned (Blais & Jungdahl, 2019). AI can be used to augment the IC's intelligence efforts and not replace it, thereby supporting and not determining the decision-making process to enable them to make better decisions. The human element ultimately should remain supreme even if it's just for cognitive and critical thinking purposes that can help to process nuances as no machine can...yet. Bias and discrimination in developing AI systems pose challenges to producing accurate results, and countering this would require highly developed processes and specialist expertise (U.K. Government, n.d.). We must maintain an active and engaged human dimension that is capable of processing nuances of the ever-changing, asymmetrical, network-centric world (Faunt & Gentile, 2019).

Some of the positive business implications for using AI responsibly are improved decision-making and customer service, stimulation of new ideas, automation, creation of new opportunities in the marketplace, and gaining valuable insights by analyzing massive volumes of data thereby enabling these organizations to make better-informed forecasts. AI can process complex data thereby giving organizations better insights to enhance efficient business practices (Martinovic, Bandur, & Tusevski, 2024). As AI can collect vast amounts of personal data which may raise ethical and privacy issues, businesses should mitigate the threat that cybercrime poses to protect such sensitive personal information.

CyberCrime and Cyber Warfare Using AI

Deception Using Deepfakes

The threat posed by real-time AI-manipulated media endangers both businesses and nation-states. Imagine taking a Zoom call from your CEO, asking you to transfer money to address an urgent business matter, but later finding out it was a cybercriminal using both video and voice deepfake technology in real-time to steal from your organization. Zror (2023) used this technology live, on-stage at the international hacker conference DEFCON, to impersonate the conference's founder Jeff Moss and state that "DEFCON is canceled" in front of nearly 30,000 attendees.

Ukrainians able to access the web in March 2022 saw a video clip of President Volodymyr Zelensky, speaking behind a podium with the Ukrainian state seal behind him in his usual battle fatigues, making a call to all the Ukrainian soldiers to lay down their arms and to return to their homes (The Guardian, 2022), as shown in Figure 1.

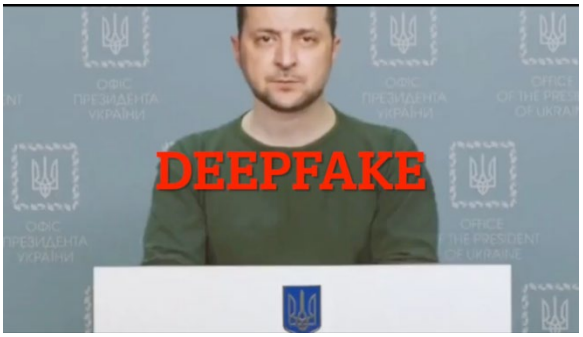

Figure 1: A deepfake video of Ukrainian President Volodymyr Zelensky appearing to urge Ukrainians to surrender to the Russian invasion in early 2022. Mikael Thalen (@MikaelThalen) Twitter, March 16, 2022, https://twitter.com/MikaelThalen/status/1504123674516885507.

Such a sophisticated hoax can have serious national security implications if taken as authentic. While sources are unsure whether the deepfake video of Zelensky released shortly after the Russian invasion of Ukraine originated in Russia, the message's intent seemed clear in asking Ukrainians to surrender to the Russian invasion. Fortunately for Ukraine, Zelensky and his team responded to and successfully debunked the deepfake within 24 hours of its appearance in world media (Simonite, 2022).

Manipulating multimedia has been made possible through the progress in machine learning, AI, and deep learning that has produced new tools and techniques. Today we have powerful tools such as the generative adversarial network (GAN) framework that assists with the production of high-resolution photorealistic videos and images. In 2019 Israel arrested three Franco-Israeli conmen who impersonated the French Foreign Minister after defrauding a businessman out of eight million Euros (The Guardian, 2019). GAN could be applied to image processing, image translation, and video synthesis (Liu, Huang, Yu, Wang & Mallya, 2021), and it should be noted that people have been blackmailed, harassed, and political discord and hate have been incited

through realistic fake and high-quality images, videos, and audios.

Deepfakes are manifested as high-quality, manipulated videos, a product of machine-learning applications that create a fake video that otherwise appears authentic by combining, replacing, merging, and superimposing video clips and images onto a video (Maras & Alexandrou, 2019). Rana, Nobi, Murali & Sung (2022) identified various approaches to tackle the challenges brought on through deepfakes and categorized them as: (1) deep learning-based techniques, (2) classical machine learning-based methods, (3) statistical techniques, and (4) blockchain-based techniques. They conclude that deep learning-based methods are by far the best in tackling the challenges deepfakes pose to decision-makers. In the end, as technology improves, efforts will need to be made to identify and expose deepfakes by developing parallel technologies to detect them.

ChatGPT as an Instrument of Misinformation

The large language model (LLM) known as ChatGPT has been making headlines since November 2022 due to the release of an advanced version capable of producing highly realistic AI-generated text that sounds like it could have been written by a human. Many cases nationwide cited the use of ChatGPT in producing homework essays that could fool college professors, and it has even been able to write legal briefs capable of fooling at least one lawyer, who is now facing sanctions for using the AI system to generate fake case law (Maruf, 2023).

In two additional cases we see that it is not a question of what information AI reveals, but more a question of what it is algorithmically programmed not to reveal. One such example ties in with Russia's control over the media, prohibiting any negative information being published about their leader, Putin. Another example is what the Chinese Communist Party allows LLMs to publish around the persecution of Uyghurs and the Tiananmen massacre (Urman & Makhortykh, 2023). We also need to be aware of the potential discrimination and biases that may be reflected in the vast data that could contribute to the ethical challenges that are associated with ChatGPT. Propagating and generating misleading information is an ethical concern (Kareem, 2024).

ChatGPT has been used to develop social engineering phishing emails designed to trick people into revealing their banking or other sensitive information, and related technologies have been used to harass individuals and endanger lives by "swatting" or calling police SWAT tactical teams with false information about non-existent threats at real addresses and homes (Cox, 2023).

As an example of using ChatGPT to generate fake news stories for major media outlets, the authors envisioned a scenario in which China might reassert its physical control over Taiwan and asked ChatGPT to write up the event in a Reuters-styled press release. The result was a believable, AI-fabricated news story that began as follows:

> *Taipei, Taiwan - Taiwan experienced an island-wide power failure on Tuesday, with experts suspecting that China used a cyber attack to black out the island. The attack left millions of Taiwanese residents in the dark and created chaos on the island. On Wednesday morning, China landed several dozen ships with 8 million civilians, doctors, engineers, and possibly plainclothes soldiers in what it labeled as a humanitarian effort to help its*

> *"little brother". The arriving ships carried food, water, and supplies, with no weapons, leaving other nations unable to intervene. While the move was initially seen as a helpful gesture, some experts have referred to it as a "bloodless coup," with China reasserting its ownership of Taiwan while causing zero casualties...*

The press release ended with the following statements, again completely AI-generated:

> *...Many fear that the situation could escalate further, raising tensions between China and the United States, and potentially leading to a larger conflict in the region. As of now, it remains to be seen what the future holds for Taiwan and its relationship with China. The situation is still unfolding, and many are watching closely to see how events will unfold in the coming days and weeks.*

Perhaps a larger concern is the speed with which such stories can be generated, in a variety of formats. In less than 5 seconds, the same story was reformatted as a Presidential Daily Brief (or PDB), a format commonly used in the US Intelligence Community. A screenshot of ChatGPT's AI-generated presidential briefing is shown in Figure 2.

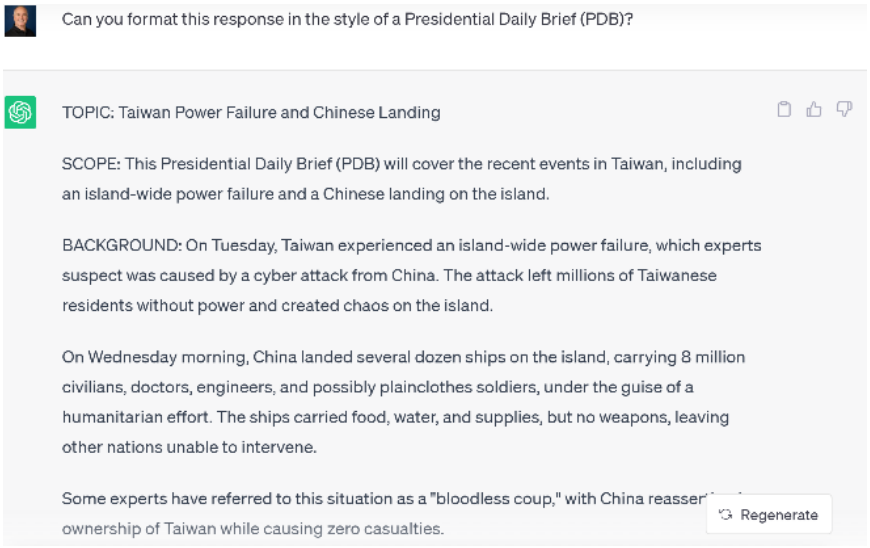

Figure 2: A fake press release from ChatGPT describing the takeover of Taiwan by China after an apparent cyberattack leading to an island-wide electrical grid blackout, formatted as a Presidential Daily Brief (PDB), a high-level US intelligence format used in the White House.

The PDB version used slightly different wording, with more detail but less prose, along with recommendations for action at the end. Using similar prompts, a small team of information warfare or psychological operations special operatives within any nation could generate literally hundreds of unique but corroborating accounts of a fictitious scenario of a similar or greater magnitude and post them to news sites or individual social media accounts, lending credibility and creating confusion regionally or world-wide.

The Generative AI Threat

An emerging threat to cybersecurity is the use of generative AI by hackers. Attackers generate fake videos, audio, images, and text by using generative AI, which allows them to launch cyberattacks such as social engineering, phishing scams, password cracking, impersonation attacks, malware development, and others. In 2019, criminals impersonated a chief executive's voice and demanded a fraudulent $243,000 transfer using AI-based voice-spoofing attack software (The Wall Street Journal, 2019), and a mayoral candidate in Chicago was impersonated by AI audio online making inflammatory remarks (Kahn, 2023), demonstrating the potential for future election misinformation to be spread using AI-generated content.

Another recent example of generative AI was used against 2024 US presidential candidate Donald Trump (Lu, 2023). The Midjourney AI image generator was used by journalist Eliot Higgins to depict Mr. Trump being arrested, at a time when Trump was being indicted on federal charges. Higgins indicated when posting the images on Twitter that he had generated the pictures using AI, but some of the pictures were captured by foreign media and presented without the information that the images were false and AI-generated (Di Placido, 2023). The image of a leading candidate from a major political party being tackled and dragged away by police (Figure 3a) is an extreme example of the kind of misinformation that generative AI can produce given even a simple query string. A more positive but still improbable image was generated by the authors by typing the query phrase "Donald Trump and Joe Biden holding hands in victory on a campaign stage with American flags," into the generative AI website, Stablediffusionweb.com (Figure 3b).

a)

b)

Figure 3a. Midjourney generative AI produced realistic-looking pictures of US presidential candidate Donald Trump being arrested, at a time when the candidate was being indicted for federal crimes. (Source: https://twitter.com/EliotHiggins/status/1637927681734987777).
Figure 3b: Stable Diffusion AI generated this slightly more light-hearted disinformation with the query "Donald Trump and Joe Biden holding hands in victory on a campaign stage with American flags." (Source: stablediffusionweb.com)

Hackers have the option of using different types of generative AI, which generates new data or

multimedia content that is difficult to distinguish from authentic human-sourced text, images, audio, or video. Present generative AI systems are typically based on generative adversarial networks (GANs), variational autoencoders (VAEs), and recurrent neural networks (RNNs) (Biniyaz, 2023). Microsoft co-founder, Bill Gates, acknowledges that AI has the potential to impact society in very profound ways, while Elon Musk, Tesla CEO, has said that AI is "actually far more dangerous than nukes" (Clifford, 2019). According to technology and AI experts, the dangers of AI, if used nefariously in digital, political, and physical attacks, could include analysis of human behaviors, speech synthesis for impersonation, moods and beliefs for manipulation, and physical weapons such as micro-drones and other physical weapons (Clifford, 2018, 2019).

Additional concerns include threats to intellectual property/copyright, security, privacy, and considerations of discrimination and bias in LLMs' output. Scams and misinformation efforts may arise because of the illegal exploitation of gen AI. AI prizes plausibility over accuracy as was shown in a 2023 court filing by a Georgia radio host against ChatGPT who falsely stated that he embezzled funds from another organization (Isik, Joshi, & Goutas, 2024). Misrepresentation by a third party of another's product can question the credibility of that product as was the case of a Tesla cybertruck crash as portrayed in a deepfake video (Isik, Joshi, & Goutas, 2024). Unaware of inauthentic content, users may inadvertently share that content, which could negatively impact the shareholder value of an organization (Isik, Joshi, & Goutas, 2024).

Application of A.I. in Information Operations

Covert action is used to achieve foreign policy objectives by influencing the way the target audience thinks and believes. Such action is sometimes referred to as information operations, propaganda, or psychological operations. By way of an example, the Soviet Union launched a fake news campaign near the end of the 20th century by spreading false rumors that the Acquired Immune Deficiency Syndrome (AIDS) disease was a US biological weapon (White House, 2022). We know that the USSR used "dezinformatsiya" (disinformation) during the Cold War to further its foreign policy objectives by sowing division and fear among its target audience. In this case, they used India's news media to plant the fake story. Had the USSR had access to AI at that time in the 1980s, this fake story would have instantly permeated social media platforms in addition to the traditional media outlets worldwide, and the intended effect of discrediting the US could have been perceived to be significantly more plausible. This fake story, at least initially, may have enjoyed credibility for a longer time causing the US to increase its efforts of discrediting the fake story, thereby distracting the US from more pressing issues of national security.

In another example of covert action at work on the political front, shortly after World War II, the Italian Communist Party threatened to oust the Christian Democratic government. To counter the surge of money coming from foreign sponsors in support of the Communists, the US had to raise funds among the anti-communist labor unions and the Italian American community to counter those efforts. Through a letter-writing campaign launched by the US government via Italian Americans to their fellow Italians, an effort was made to convince the Italians that life under a capitalist system is much better than under a communist system (Acuff et al., 2022). Had AI existed at that time, this information operation could have reached a much wider audience in

nanoseconds via various social media platforms and resulted in a significant impact on the voters not to support the Communist Party.

Another area of covert action is the removal of a dangerous foreign leader through the fomenting of a *coup d'état* in place of waging war against that country. An example of such covert support to a *coup* was the removal of the socialist president of Chile, Salvador Allende, in the early 1970s. Unfortunately, he was replaced by Augusto Pinochet, who ran a brutal military regime until 1990. The CIA denied any involvement in the coup, although it is widely believed that the US Government was behind the plot (Acuff et al., 2022). Had AI existed at that time, it would have been easier for the US government to proffer plausible deniability by waging an information campaign via social media and other media platforms aimed at discrediting those rumors and allegations in real-time.

Sabotage does not only manifest as violence, but it could also be non-violent in nature if it is intended to foment instability through information operations. An example of such non-violent means was Russian covert acts to sow dissension and confusion among Americans when it was aimed at Hillary Clinton's presidential bid in 2016 (Acuff et al., 2022). If AI had been developed where deepfakes could have been used in the Russian covert operations, it would have made their operation more plausible and effective by impersonating her and other opposition candidates' fake responses to enhance even more division among the populace.

Today, Russia appears to be embracing a new model of sabotage referred to as the so-called gig economy, one which focuses on a temporary workforce made up of freelance contractors in a largely free-market online system to support Putin's plausible deniability efforts when nefarious cyber operations are conducted against enemy states (Richterova et al., 2024). It appears that Russia's current sabotage operations mostly resemble the crowdsourced type, used for recruitment on Telegram by Russian intelligence services (Richterova et al., 2024).

Covertly facilitating strikes that target a large company, an industry, or a country's economy would be another example of sabotage complemented by a disinformation campaign that could undermine investor and consumer confidence. Having access to AI platforms can enhance the messaging to reach a wider intended audience in a much shorter time thereby amplifying the intended effect of undermining the economy. Fake imagery, audio, and video of major economic and political players, along with a tidal wave of fake news stories, social media posts, and even disinformation spread by AI influencers—fake or licensed personas where all content is generated by AI—can, and likely will, be combined into a perfect storm of disinformation, distraction, and disruption.

Furthermore, cyber financial sabotage against a nation or large corporation can potentially deter investment, disrupt economic activities, and undermine confidence in the stability of the company or financial system. Currency manipulation and economic sanctions may lead to capital flight, market volatility, and recessionary pressures with their concomitant negative effects on growth and economic instability, which may exacerbate poverty and social inequality (Green, 2024).

Responding to Real-Time Disinformation: A Brief Case Study

How, then, can businesses, organizations, and governments plan for and respond to the emerging threat of real-time information warfare from AI-enhanced adversaries including nation-states, terrorist groups, organized crime, and rogue individuals? The Zelensky case study offers a tested approach: developing a playbook for disinformation incident response. This playbook is similar to cybersecurity incident response plans developed over the past three to four decades for businesses and governments, plans now considered crucial for compliance and business continuity/disaster recovery planning.

Ukraine's swift response to the Zelensky deepfake demonstrates some of the top priorities in responding to disinformation from AI and can serve as an exemplar for governments, organizations, and individuals alike. First, the response was planned. Ukraine was prepared for potential Russian disinformation and had developed a playbook in advance for responding to deception in near-real-time. Information security incident response plans are required in most midsize to large organizations so that IT and other staff can respond quickly and effectively to cyber breaches, ransomware attacks, and similar events. Planning in advance for disinformation incidents against government agencies or officials, businesses, and individuals may be seen as a common compliance obligation in a similar fashion in the foreseeable future.

Further, the response to the alleged Russian deepfake video was immediate and took advantage of both traditional and online media. Within minutes of the fake video's appearance on television, President Zelensky posted a personal response via Facebook video repudiating the video's message and discrediting the deepfake (Simonite, 2022). News sources cited the speed and authenticity of the response as critical in dealing with the deepfake video and dispelling rumors before they could take root.

The video response was a live recording of the President, personalized and authentic to the situation, and it directly refuted the deepfake video's message and claim. Media outlets covered the incident, along with Zelensky's response. This could be applied to any organization's disinformation playbook by preparing public relations, social media, marketing, or similar staff to capture and disseminate authentic videos of the CEO or other targeted individuals addressing a deepfake video or false, AI-generated disinformation head-on.

In the Zelensky case, Facebook and YouTube eventually deleted uploads of the deepfake video, as deceptive or manipulated media is a violation of their terms of service, but it was Zelensky's quick, thorough, personal, and authentic response, combined with his public relations team's swift distribution of that message through both social media and traditional media outlets that neutralized the potentially harmful situation before it had a chance to be disseminated widely (Allen et al., 2023).

Developing a Disinformation Incident Response Playbook

Based on the Zelensky case study and preceding literature review, below are the authors' recommendations for governments, organizations, and individuals as they develop playbooks for responding to AI-generated or AI-assisted disinformation in near-real-time:

1. **Be prepared**. Develop a playbook or have a plan in advance for responding to disinformation events, just like your organization is likely required to do for cybersecurity incidents. Larger organizations should include handling disinformation via social media in their table-top preparedness exercises alongside natural disasters and cyber incidents for business continuity and disaster recovery scenarios.

2. **Respond quickly, personally, and authentically**, preferably with live video and audio of the subject of the disinformation. Zelensky's immediate response, in the form of a personalized smartphone video that was distributed first on Facebook and then through news outlets, was key to halting Russia's alleged disinformation operation. If a CEO, world leader, or individual is the subject of a deepfake, digital voice impersonation, fake images, misleading AI-generated online news articles, or other disinformation, that person must be ready to respond in near-real-time to remove any momentum from the fake media before it spreads beyond containment.

3. **Use traditional media and online, social media** to refute disinformation across all platforms. Beginning with the platform where the original disinformation was shared, which was Facebook in the Zelensky case, was a crucial step in Ukraine's response, but they didn't stop there. Zelensky's team quickly called in the aid of the nation's press and then world news organizations in curtailing the spread of the falsified deepfake video. Depending on the situation and the severity, posting a video to multiple social media platforms would be followed by calling a press conference or reaching out to local and national news outlets. In most cases, multiple forms of media should be engaged as part of an organization's playbook when fighting information warfare.

4. **Follow through** with news outlets and online sites to remove manipulated and deceptive media to ensure that it isn't re-disseminated later. In the Zelensky case, Facebook and YouTube removed the altered videos within hours, and news organizations superimposed the word "DEEPFAKE" on videos and still-frames featuring the manipulated media to help ensure it would not be reposted and misinterpreted as authentic. Sites like Facebook, X (formerly Twitter), YouTube, and other social media giants may take more time to remove posted videos or audio recordings, but they are a valuable part of the process. And include instructions in your organization's disinformation incident response playbook for your own public relations team to indelibly mark all falsified video, audio, and images as "DEEPFAKE", "FALSE INFORMATION", or "AI-GENERATED" so that they will be readily identifiable as fake, even when reposted or shared out-of-context.

5. **Train employees** to be suspicious of and verify not only email and text messages, but also phone and video calls. In every social engineering attack, awareness is a key factor. Social engineering awareness is probably already a part of your organization's cybersecurity incident response plan, but highlight the threat of deepfakes and AI-generated disinformation so your team can recognize and respond to attacks as quickly and effectively as the Ukrainian leadership team. Make your employees and top-level executives aware of the danger of deepfake voice and video calls, as well as AI-generated emails, text messages, web pages, and news articles. Educate and

empower employees to verify all requests before complying.

Testing the Playbook

How Organizations Can Respond to AI-Generated Fake News Postings

Let's apply the recommendations in the preceding section to the case of disinformation via massive fake news postings like the one about China's takeover of Taiwan presented earlier. First, business leaders and governments of both China and Taiwan should be prepared for disinformation events such as this one. Communications professionals in each government would need to develop a rapid repudiation message, using live video of both television personalities and government officials, preferably live in recognizable, public spaces in Taiwan, showing that there was no actual invasion. Such messages would need to be posted online and broadly disseminated via traditional news and media outlets. For Taiwan, the goal would be to prevent disruption of the economy and trade with peer nations. For China, it would be equally important to prevent international sentiment from turning against China and avoid sanctions, as well as other negative outcomes.

Responding to False Generative AI Images, Video, and Audio

Politicians, CEOs, organizations, and governments must consider the future need to counter disinformation in their planning exercises. The false images of the Trump arrest were noted by the original poster to be fictitious, just as the image of Clinton and Trump kissing was presented as AI-generated by the authors of this manuscript. But if similar, or even more damaging, images, video, and/or audio were released by a rival, an adversary, malicious government, criminals, or terrorists, both candidates and leaders alike would need to be prepared to respond immediately, personally, and authentically via live video to address the specific disinformation being spread.

The Ultimate Challenge: Responding to Real-Time Deepfakes and Disinformation

Experts across industries have been predicting for years that technology would advance to the point that real-time deepfake video, audio and similar disinformation using AI would be possible, and at the DEFCON hacking conference late last year, those fears were found to have come true (Zror, 2023). Security researcher Gal Zror demonstrated a combination of already-existing technologies chained together to replace his face and voice in real-time with the founder of DEFCON, Jeff Moss, with only a slight delay similar to what we have come to expect from long-distance web conferences or satellite correspondents' video during a live news broadcast. Zror jokingly misinformed the crowd that the DEFCON conference was canceled (Zror, 2023).

Responding to a malicious actor who has stolen both a leader's face and voice, and who can both speak with news agencies and post live videos to social media and online sources, may be the pinnacle of dispelling disinformation. In addition to all the components discussed above in forming a playbook for responding to disinformation, maintaining a personal relationship with multiple media outlets and personalities, or at least with a well-connected public relations firm, would help mitigate the damage such an actor could inflict. It could very well be a near-term need for an affected politician or business leader to personally call a news anchor on their mobile

phone to show that they're the real person—but a highly resourced malicious actor could easily spoof the phone number of the celebrity or leader they're impersonating.

All elements of the playbook would need to be deployed quickly to be able to respond to publicly posted videos, but the final playbook recommendation, training our employees to be aware of the threat of deepfake social engineering, would be crucial to stopping targeted fraud and theft like this.

Beyond our own organizations, educating members of the media, and the general public, as well as ourselves, our friends, and family, of the present danger posed by AI-enhanced impersonators must become part of the strategy for dispelling disinformation before it spreads. The old-fashioned journalistic tenet of verifying a source and the information provided by that source has become both more difficult and more vital than ever. But as traditional media sources fade into the background amidst the clamor of misinformation inadvertently shared and re-posted by our own friends and online connections, the prospect of cutting off disinformation before it spreads becomes a matter of incident response and business continuity planning.

Conclusions

While the world has not yet experienced, as of the time of this writing and to the best of our knowledge, a large-scale information operation leveraging the full extent of deepfake misinformation, or fake news, audio, video, or images produced by generative AI, the near-term prospect has been demonstrated and is a current threat to leaders of nations, for-profit and non-profit organizations, and private individuals. In the not-too-distant future, a video teleconference with your CEO asking for an urgent wire transfer, a phone call with your nephew asking for money, or a public address by the leader of your nation claiming to have just stepped down from power may be realistic enough to fool the majority, or at least enough to have the intended effect of defrauding, distracting, disrupting, or worse.

This research extends and develops the recommendations of previous authors through the lens of the case study on Ukraine's response to alleged Russian disinformation operations against Ukrainian President Volodymyr Zelensky near the outset of the Russian invasion of Ukraine. The goal of this research is to encourage organizations and leaders to prepare a disinformation incident response playbook to respond to potentially damaging AI-generated content. This playbook approach advocates for targeted organizations and individuals to respond in real-time, using personalized, authentic messaging to dispel fake video, audio, or images. Furthermore, the target of a disinformation campaign should use both traditional media outlets and online/social media platforms together in repudiating misleading video/audio/images/text, and the affected individual or organization must follow up with online platforms after the fact to ensure that manipulated or AI-generated false information is taken down to avoid further redistribution.

Ultimately, citizens, investors, and leaders will have to rely either upon more advanced detection systems (algorithms driven by more AI) or upon better education for journalists, organizations, ourselves, and the public at large in verifying sources and information before sharing. Organizations and governments can plan to successfully mitigate the reputational damage and real-world destruction that can be wrought by ChatGPT, generative AI in general, and deepfake-manipulated media beginning with the playbook uncovered in this research.

In the meantime, we are left to contend with the fact that we simply cannot believe what we see or hear in audio, video, images, or text until it is checked, rechecked, and verified—something the intelligence community has known for decades, but now at a scale and velocity that can impact economic and global stability faster than a news cycle. But by developing a playbook for dealing with AI-generated or AI-enhanced disinformation ahead of time, we can enhance the effectiveness of our incident response and improve the odds of our businesses' or governments' survival from such foreseeable, tangible, and near-term threats.

References

Allen, C., Payne, B.R., Abegaz, T.T., Robertson, C.L. (2023). What You See Is Not What You Know: Studying Deception in Deepfake Video Manipulation. *Journal of Cybersecurity Research, Education, and Practice (JCERP)*, 2024 (1), Article 1, 7 pp. October 2023. ISSN 2472-2707. https://digitalcommons.kennesaw.edu/jcerp/vol2024/iss1/1/

Australian Strategic Policy Institute (ASPI). (2023). *ASPI's Critical Technology Tracker: The global race for future power* (Policy Brief Report No. 69/2023). ISSN 2209-9670. Retrieved from https://www.aspi.org.au/report/critical-technology-tracker

Biniyaz, J. (2023, May 17). Generative AI: Posing Risk of Criminal Abuse. *Readwrite.* Retrieved from https://readwrite.com/generative-ai-posing-risk-of-criminal-abuse/.

Bipartisan Policy Center (2020). *Artificial Intelligence and National Security.* June 2020. Retrieved from https://bipartisanpolicy.org/report/ai-national-security/

Blais, J. R., & Jungdahl, A. M. (2019). Artificial Intelligence in a Human Intelligence World. *American Intelligence Journal*, *36*(1), 108–113.

Clifford, C. (2019, March 26). Bill Gates: A.I. is like nuclear energy — 'both promising and dangerous'. *CNBC.* Retrieved from https://www.cnbc.com/2019/03/26/bill-gates-artificial-intelligence-both-promising-and-dangerous.html

Clifford, C. (2018, February 21). Top A.I. experts warn of a 'Black Mirror'-esque future with swarms of micro-drones and autonomous weapons. *CNBC*. Retrieved from https://www.cnbc.com/2018/02/21/openai-oxford-and-cambridge-ai-experts-warn-of-autonomous-weapons.html

Cox, J. (2023, April 13). A Computer Generated Swatting Service Is Causing Havoc Across America. *Vice.com*. Retrieved from https://www.vice.com/en/article/k7z8be/torswats-computer-generated-ai-voice-swatting

Defense Advanced Research Projects Agency (DARPA). (2018, September 7). *DARPA Announces $2 Billion Campaign to Develop Next Wave of AI Technologies*. Retrieved from https://www.darpa.mil/news-events/2018-09-07

Di Placido, D. (2023, March 22). AI-Generated Images of Donald Trump Getting Arrested Foreshadow a Flood of Memes, Fake News. *Forbes.* Retrieved from https://www.forbes.com/sites/danidiplacido/2023/03/22/ai-generated-images-of-donald-trump-getting-arrested-foreshadow-a-flood-of-memes-fake-news/

Farah, H. (2023, August 2). Humans can detect deepfake speech only 73% of the time, study finds. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2023/aug/02/humans-can-detect-deepfake-speech-only-73-of-the-time-study-finds

Faunt, R. A., & Gentile, P. D. (2019). Artificial Intelligence within the Intelligence Community: The Need to Retain the Human Dimension. *American Intelligence Journal*, *36*(2), 48–53.

Green, J. I. A. S. W. B. (2024). Unraveling the Dynamics and Impacts of Financial Sabotage: A Comprehensive Analysis. *Researchgate.* Retrieved from https://www.researchgate.net/ profile/ Jamell-Samuels/publication/378653936_Maternal_Haplogroups_and_ Performance_in_Long-Distance_Running_A_Comprehensive_Analysis/

Isik, O., Joshi, A., & Goutas, L. (2024, May 31). 4 Types of Gen AI Risk and How to Mitigate Them. *Harvard Business Review.* Retrieved from https://hbr.org/2024/05/4-types-of-gen-ai-risk-and-how-to-mitigate-them

Kareem, T. A. (2024). Impact of Chat GPT on Human Communication and Social Interaction. *International Journal of Engineering and Modern Technology*, *10* (8), pp. 30-50. Retrieved from https://iiardjournals.org/

Liu, M. -Y., Huang, X., Yu, J., Wang, T.-C, & Mallya, A. (2021). Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications. In *Proceedings of the IEEE, 109*(5), pp. 839-862.

Lu, D. (2023, March 31). Misinformation, mistakes and the Pope in a puffer: what rapidly evolving AI can – and can't – do. *The Guardian.* Retrieved from https://www.theguardian.com/ technology/2023/apr/01/misinformation-mistakes-and-the-pope-in-a-puffer-what-rapidly-evolving-ai-can-and-cant-do

Kahn, J. (2023). Fake News 2.0: The Election Threat from A.I. *Fortune*. Retrieved from https://fortune.com/2023/04/08/ai-chatgpt-dalle-voice-cloning-2024-us-presidential-election-misinformation/

Mai, K. T., Bray, S., Davies, T., Griffin, L. D. (2023, August 2). Warning: Humans cannot reliably detect speech deepfakes. *PLoS ONE 18*(8): e0285333. Retrieved from https://doi.org/10.1371/journal.pone.0285333

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof, 23*(3), pp. 255-262.

Maruf, R. (2023, May 28). Lawyer apologizes for fake court citations from ChatGPT. CNN. Retrieved from https://edition.cnn.com/2023/05/27/business/chat-gpt-avianca-mata-lawyers/index.html

Martinovic, G., Bandur, K. M., & Tusevski, B. (2024, April). The Impact of Artificial Intelligence on Organizational Structure Trends Analysis and Implications. In *Economic and Social Development (Book of Proceedings), 110th International Scientific Conference on Economic and Social* (Vol. 5, p. 149).

Milligan, S. (2023, May 23). White House Pledges 'Road Map' for Managing Artificial Intelligence. *U.S. News.* Retrieved from https://www.usnews.com/news/politics/articles/2023-05-23/white-house-pledges-road-map-for-managing-artificial-intelligence

Moran, C.R., Burton, J., Christou, G. (2023). The US Intelligence Community, Global Security, and AI: From Secret Intelligence to Smart Spying. *Journal of Global Security Studies*, *8*(2), 2023.

Rana, M. S., Nobi, M.N., Murali, B., & Sung, A.H. (2022). *Deepfake Detection: A Systematic Literature Review. IEEE Access 10*, pp. 25494-25513. Retrieved from https://ieeexplore.ieee.org/abstract/document/9721302

Richterova, D., Grossfeld, E., Long, M., & Bury, P. (2024). Russian Sabotage in the Gig-Economy Era. *The RUSI Journal*, *169*(5), pp. 10–21. Retrieved from https://doi.org/10.1080/03071847.2024.2401232

Shapiro, C. (2022, October 4). The Intelligence Community Is Developing New Uses for AI. *FedTech.* Retrieved from https://fedtechmagazine.com/article/2022/10/intelligence-community-developing-new-uses-ai-perfcon

Simonite, T. (2022, March 17). A Zelensky Deepfake Was Quickly Defeated. The Next One Might Not Be. *Wired*. Retrieved from https://www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/

Stupp, C. (2019, August 30). Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal.* Retrieved from https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

The Guardian (2022, March 19). Deepfakes v pre-bunking: is Russia losing the infowar? Retrieved from https://www.theguardian.com/world/2022/mar/19/russia-ukraine-infowar-deepfakes

The Guardian (2019, March 29). Conmen made €8m by impersonating French minister - Israeli police. Retrieved from https://www.theguardian.com/world/2019/mar/28/conmen-made-8m-by-impersonating-french-minister-israeli-police

The White House. (2022). *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. Retrieved from https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Tucker, P. (2020, January 27). Spies Like AI: The Future of Artificial Intelligence for the US Intelligence Community. *Defense One*. Retrieved from https://www.defenseone.com/technology/2020/01/spies-ai-future-artificial-intelligence-us-intelligence-community/162673/

Urman, A., & Makhortykh, M. (2023, September 8). The Silence of the LLMs: Cross-Lingual Analysis of Political Bias and False Information Prevalence in ChatGPT, Google Bard, and Bing Chat. *OSFPREPRINTS*. Retrieved from https://doi.org/10.31219/osf.io/q9v8f

United Kingdom Government (n.d.). *Pioneering a new National Security. The Ethics of Artificial Intelligence.* Retrieved from https://www.gchq.gov.uk/artificial-intelligence/accessible-version.html

Zror, G. (2023). Look Ma, I'm the CEO: Real-Time Video and Audio Deepfakes. *DEFCON 31.* Retrieved from https://media.defcon.org/DEF%20CON%2031/DEF%20CON%2031 presentations/

Quarterly Review of Business Disciplines

Volume 11 Number 3/4 February 2025

This issue is now available online at www.iabd.org.

A Journal of International Academy of Business Disciplines
Sponsored by University of North Florida
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)