Fordham Law School

FLASH: The Fordham Law Archive of Scholarship and History

2019

From the Myth of Babel to Google Translate: Confronting Malicious Use of Artificial Intelligence— Copyright and Algorithmic Biases in Online Translation Systems

Shlomit Yanisky-Ravid
Fordham University School of Law, syaniskyravid@fordham.edu

Cynthia Martens
Deborah A. Nilson & Associates, PLLC

Follow this and additional works at: https://ir.lawnet.fordham.edu/faculty_scholarship

From the Myth of Babel to Google Translate: Confronting Malicious Use of Artificial Intelligence—Copyright and Algorithmic Biases in Online Translation Systems

Professor Shlomit Yanisky-Ravid and Cynthia Martens*

Many of us rely on Google Translate and other Artificial Intelligence and Machine Learning (AI) online translation daily for personal or commercial use. These AI systems have become ubiquitous and are poised to revolutionize human communication across the globe. Promising increased fluency across cultures by breaking down linguistic barriers and promoting cross-cultural relationships in a way that many civilizations have historically sought and struggled to achieve, AI translation affords users the means to turn any text—from phrases to books—into cognizable expression.

This Article discusses the burgeoning possibilities in the 3A Era (Advanced, Autonomous, AI systems) of AI online translation as accessible tools, whose users are data suppliers and feedback providers, and hence, contributors to the programing and improvement processes of these translation tools. On the other hand, this Article also acknowledges the real concerns this new realm raises, stemming from malicious uses of AI, which are most often concealed from the public. Such hidden aspects include built-in algorithmic biases, such as race, sex and gender, color, religion, or national origin biases, which this Article addresses in a discussion of AI systems' systemic shortcomings. Because AI translation systems learn and function through the data that they receive from data providers, they are vulnerable to societal biases. When users offer

* Shlomit Yanisky-Ravid Professor of Law, PhD. Fordham Law School, Visiting Professor, Fordham CLIP, Head of the IP-AI & Blockchain Project. Yale Law School, ISP, Professor Fellow. Ono Law School (OAC), Israel, Associated Professor of Law, The Shalom Comparative Legal Research Institute, Founder and Academic Director.
Cynthia Martens, Fordham University School of Law, J.D. 2019. CLIP IP-AI & Blockchain Project Fellow. Professional journalist and translator. Associate at Deborah A. Nilson & Associates, PLLC (admission to the New York bar pending).

feedback, these systems may perpetuate sexist, racist, or otherwise objectionable expressions, of which other users, when consulting the systems, are unaware.

Furthermore, examining the current copyright regime, this Article claims, for the first time, that we, as users, have become inadvertent infringers of legal rights since a translation, according to copyright law, is a derivative work owned and controlled by the author. As such, an author's permission is necessary prior to the creation of a translation, with the author in a position to collect payment when due. Moreover, under current law, the Fair Use Doctrine is frequently inapplicable. This Article claims that the legal and academic communities and policy makers have failed to address AI translation systems within the copyright regime, and that this failure renders the current copyright regime outdated and ill-equipped to handle the advent of sophisticated AI tools. Additionally, this Article states that the present inability of AI technology to routinely capture the nuance of human prose gives rise to another concern. The ubiquitous role such (as yet) flawed AI online translation systems play in translation services, for personal or commercial purposes, should be better balanced with the concerns of authors—who may worry about the linguistic integrity of an AI translation of their work—and their rights, in certain circumstances, to control translations of their work and object to unauthorized AI translations.

Understanding the concerns attending unauthorized AI translations under the current copyright regime, while still recognizing that users should be able to profit from the wellspring of literacy which AI translation offers, this Article argues for a harmonization of AI translation with amended copyright protection.

To that end, this Article calls on policymakers to adjust the current legal regime to include advanced technologies and suggests new principles for combining legal tools with existing technology. Such an approach would better balance the benefits of accessible AI translation systems with the requirements of a modified, modern copyright regime via the implementation of a method coined "fair use and equality by design." Additionally, by recognizing the conflicting interests at stake, this Article invites international policymakers, such as WIPO (the World Intellectual Property Organization), to promote the development of international standardized guidelines for the use of AI translation systems, and possibly other AI systems, by emphasizing fair use exemptions and limitations.

This Article concludes that by understanding the significant drawbacks of AI translation systems and adopting the suggested principles, policymakers can promote access to an evolving AI technology, while also recognizing the integrity of authors' linguistic

choices and preserving the beauty of linguistic diversity—which, as the ancient story of Babel hinted, is valuable because of, not despite, the challenges it poses.

CONTENTS

I. INTRODUCTION

Linguistic barriers have frustrated people for thousands of years. In the famous story of the Tower of Babel, Genesis 11:1–9 describes how "the whole earth had one language and the same words" until its denizens sought to build "a tower with its top in the heavens," and God intervened to curtail this excessive ambition, saying, "Come, let us go down, and confuse their language there, so that they will not understand one another's speech."[1] From then on, according to biblical history, the world's inhabitants were scattered across the globe, unable to understand each other's languages.[2] Recently, however, Google Translate and other web-based AI translation systems have disrupted this narrative by allowing people to quickly access translations online.[3]

Though it can be tempting to think translations require no more than replacing a word in one language with the corresponding word in another,

---

1. 7 *Genesis* 11:1–9 (New Revised Standard Version).

2. *Id.*

3. *See generally* Gideon Lewis-Kraus, *Is Translation an Art or a Math Problem?*, N.Y. TIMES (June 4, 2015), https://www.nytimes.com/2015/06/07/magazine/is-translation-an-art-or-a-math-problem.html [https://perma.cc/X8EC-DTPM].

this has never been true.[4] Translations are as powerful and complex as the languages and the advanced technology that produce them.[5]

In 2006, Google launched its online Google Translate service. Since then, this tool has become ubiquitous, as users easily and frequently turn to the world's dominant search engine to translate everything from quick phrases to lengthy texts.[6] Though the benefits and convenience of Google Translate and other similar programs are obvious, their use has caused large swathes of the global population to become inadvertent infringers of copyright law because legally *authors*, and not *users* or search engines, have the exclusive right to control translations of their works. Even the American Fair Use Doctrine does not present a consistently good defense to this use of corporate translation platforms, which do not limit the amount of works users can borrow. What is more, because such platforms make money off of users' translation queries, their use of the original works is commercial and possibly supplants the market for an authorized translation.

The question then becomes whether the traditional copyright regime can meet the challenges of Artificial Intelligence and Machine Learning (AI) online translation systems,[7] or whether this regime is outdated and in

---

4. *See* Cassandra L. McKeown & Michael G. Miller, *Say What?: South Dakota's Unsettling Indifference to Linguistic Minorities in the Courtroom*, 54 S.D. L. REV. 33, 41–42 (2009) ("Historically, courts have embraced the 'conduit theory' of interpretation. This theory views the interpreter as a machine into which one language enters and another language exits . . . . Regardless of the cause, the bench and bar regularly underestimate the difficulty of facilitating accurate communication between languages and cultures."); *see also* Stella Szantova Giordano, Note, *It's All Greek to Me: Are Attorneys Who Engage in or Procure Legal Translation for Their Clients at Risk of Committing an Ethical Violation?*, 31 QUINNIPIAC L. REV. 447, 448 (2013) ("While the legal community often assumes that legal translation is merely a simple and straightforward mechanical process, akin to an administrative function, linguists and comparative lawyers vehemently disagree. Any type of foreign language translation is an inherently imperfect and imprecise process.").

5. *See* Lewis-Kraus, *supra* note 3.

6. *Id.* ("Google Translate is far and away the venture that has done the most to realize the old science-fiction dream of serene, unrippled exchange. The search giant has made ubiquitous those little buttons, in email and on websites, that deliver instantaneous conversion between language pairs. Google says the service is used more than a billion times a day worldwide, by more than 500 million people a month.").

7. In this Article, we will use the phrases AI and ML interchangeably to refer to online translation programs. *See* Debora Person, *AI Application in the Practice of Law*, WYO. LAW., Aug. 2017, at 50 ("Artificial intelligence is an umbrella term. It refers to natural language processing and machine learning functions."). For a discussion of artificial intelligence systems in general and the intellectual property regime, see Shlomit Yanisky-Ravid, *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era—The Human-Like Workers Are Already Here—A New Model*, 4 MICH. ST. L. REV. 659 (2017) [hereinafter Yanisky-Ravid, *Generating Rembrandt*] (discussing artificial intelligence systems that produce works of art and suggesting that the work-for-hire doctrine would allow for imposing accountability and ownership on the user); *see also* Shlomit Yanisky-Ravid & Xiaoqiong (Jackie) Liu, *When Artificial Intelligence Systems Produce Inventions: An Alternative Model for Patent Law at the 3a Era*, 39 CARDOZO L. REV. 2215, 2223–24 (2018)

need of amendment or replacement. Copyright is not, however, the only issue web translation tools raise. Because online translation tools predominantly rely on AI, their end products are vulnerable to algorithmic biases users are often unaware of; widespread reliance on incorrect translations can perpetuate discrimination, facilitate rights violations, and lead users to communicate with prejudiced or inappropriate terminology.[8]

AI online translation tools are by far among the most accessible and user-friendly of AI systems, translating with just a few clicks of a mouse or taps on a smartphone screen. As advanced technology tools have developed rapidly, AI translation systems have also become available on newer devices, including, reportedly, Amazon's virtual assistant Alexa and Apple's well-known Siri.[9] Popular AI translation tools such as Google Translate, Microsoft Translator, DeepL or Systran, among many other options, transfer written materials including emails, advertisements, news, articles, songs, literature, books, and lectures from one language into another. These tools create derivative works by translating copyrighted written materials with just a quick click.[10] In many ways, by eliminating linguistic hurdles and enabling international communication, this advanced technology revives questions at the heart of the Biblical story of the Tower of Babel.

Interpretations of the story of Babel are varied. What was it about the building of the tower that warranted divine interference? Why did God's intervention consist of sowing such widespread linguistic confusion? Theories abound: some suggest that God was distressed by the dangerous potential for people using "the same words" to exhibit a uniformity of thought, rather than promoting diversity, which is beneficiary to the growth and the development of humankind.[11] Others argue that the tower

---

(describing the use of AI systems to produce creative works and the applicability of intellectual property laws, which in some ways have become irrelevant and outdated, with a focus on patent law).

8. *See infra* note 146.

9. *See, e.g.*, Dave Gershgorn, *Cultural Context Is the Hardest Part of Speaking a Foreign Language. Amazon's Alexa Could Change That*, QUARTZ (Mar. 1, 2018), https://qz.com/1219383/amazon-alexas-language-translation-ai-will-reportedly-learn-cultural-context [https://perma.cc/BS5Y-SYF5] (citing a report from Yahoo Finance suggesting that Amazon "is looking to amp up Alexa's translation skills, by also including how to respond to cultural norms").

10. *See* Lewis-Kraus, *supra* note 3.

11. Shai Held, *The Babel Story is About the Dangers of Uniformity*, CHRISTIAN CENTURY (Oct. 24, 2017), https://www.christiancentury.org/article/critical-essay/the-babel-story-is-about-dangers-uniformity [https://perma.cc/PT29-U72G]; *see also* YESHAYAHU LEIBOWITZ, SEVEN YEARS OF CONVERSATIONS OVER THE WEEKLY TORAH PORTION 1976–1982, at 30–35 (Ben Zion Mishael Nuriel ed., 2000); Andrea Cantor, *Babel: The Biblical Imperative for Diversity*, JEWISH EXPONENT (Apr. 18, 2013), http://jewishexponent.com/2013/04/18/babel-the-biblical-imperative-for-diversity [https://perma.cc/F9UX-U9FM].

symbolized mankind's "arrogant pursuit of fame and power."[12] It is also true that humanity gained tremendously from the emergence of different languages and dialects, each of which is connected to a particular place and culture. Nevertheless, to the popular imagination, Babel, a mysterious story "among the most famous in the Bible,"[13] often represents a tragic loss of communication among people—a barrier and source of divisiveness.[14] Perhaps this loss explains the impassioned promotion of AI online translation systems as revolutionary tools poised to return the world to a pre-Babel era by creating an online bridge across languages—one that enhances communication and supports peace.[15] On the other hand, in addition to transforming virtually all users into unwitting copyright infringers, these systems can also produce inaccurate and biased translations that reflect poorly on their creators and users alike, not to mention the original authors whose works new readers encounter for the first time in translated form.

This Article, therefore, discusses the challenges that AI translation tools pose to users, authors, and copyright law. As these systems have become omnipresent, insufficient attention has been paid to the copyright regime, to the need to incentivize authors while promoting the public welfare, to the importance of accurate translations, and to the right of authors to object to adulteration of their works. This Article argues that in the "3A era" (namely, that of Advanced, Automatic, and Autonomous AI systems), intellectual property policymakers should specifically address online translation to eliminate legal confusion and better balance conflicting needs. The interests at stake include those of the public, which benefits from easily accessible translation tools, of authors, who have the right to control the use of their works, and of technology firms, which currently profit from free access to both user data and author material. Since translation consists of converting one language to another, and since the use of AI translation systems has reached global proportions, so, too, are the legal challenges these systems pose in a global scope. A comparative legal awareness and analysis are therefore necessary to develop a unified solution reflective of differences in national regulation.

---

12. Jan Christian Gertz, *The Tower of Babel (Gen 11:1–9)*, BIBLE ODYSSEY (Feb. 7, 2019), http://www.bibleodyssey.com/passages/main-articles/tower-of-babel [https://perma.cc/J26R-C624].

13. Brent A. Strawn, *Holes in the Tower of Babel*, OXFORD BIBLICAL STUD. ONLINE, https://global.oup.com/obso/focus/focus_on_towerbabel (last visited Sept. 17, 2019).

14. *See* Lewis-Kraus, *supra* note 3 ("It is a dream that harks back to Genesis, of a common tongue that perfectly maps thought to world. In Scripture, this allowed for a humanity so well coordinated, so alike in its understanding, that all the world's subcontractors could agree on a time to build a tower to the heavens. Since Babel, though, even the smallest construction projects are plagued by terrible delays.").

15. *Id.*

Although many of the materials being translated online are copyrightable assets, users of AI translation tools do not generally contact rightsholders for permission, and thus have neither obtained licenses to use the works nor paid royalties for them. As a result, on an everyday basis many are violating copyright law. This Article assumes that a majority of users are not aware of the legal protections afforded to translations. Translation is explicitly defined in the Copyright Act as a derivative work, whose copyrights belongs to the author (or to an author-approved person or entity).[16] We argue that these routine infringements further weaken the validity and influence of copyright, which has been damaged by a libertarian internet philosophy that totally ignores authorship and accuracy.[17] While undoubtedly many applications of AI translation constitute fair use by individual internet users, it is unclear why wealthy technology companies such as Google, Amazon, and Microsoft should have free access to authors' derivative rights. Copyright is a balancing act that aims to protect free expression and serve the public welfare by providing creative individuals with limited-time incentives to share the fruits of their labor. By ignoring the conflicts of AI translation, lawmakers are spreading confusion—which is only likely to grow as AI develops further—and undermining copyright law, which becomes meaningless if not enforced.

We acknowledge that our arguments raise conflict between traditional copyright protections and the benefits of advanced technology—one which no one wants to confront, despite the fact that most users are technically violating current copyright law when they engage with Google Translate. However, we do not advocate for the elimination of AI translation tools, or for an overly expansive version of copyright law. Instead, we argue that the law must better reconcile the interests of the different stakeholders in order for the "Progress of Science and useful Arts"[18] to flourish for the benefit of all. We call for a re-thinking of copyright laws that stresses the importance of both AI translation tools and copyright ideals by implementing principles of proportionality and reasonableness. Given the inherently international nature of the internet

---

16. 17 U.S.C. §§ 101, 106 (2018). "A 'derivative work' is a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a 'derivative work.'" 17 U.S.C. § 101 (2018).

17. *See, e.g.*, 63 AM. JUR. 3D *Proof of Facts* § 1 (2001) (discussing copyright infringement by file sharing on the internet).

18. U.S. CONST. art. I, § 8, cl. 8.

and translation, we posit that the best solution will involve input from comparative and international lawmakers.

This Article both recognizes the value of AI translation systems and questions their unfettered use, which—as stressed—leads to copyright violations and unreliable, potentially biased language. Following the discussion of this state of affairs and its attendant concerns, we make several recommendations. First, a combination of digital and legal tools should balance the competing interests at stake. Second, a new AI translation system paradigm requires greater consideration for authors' rights in translations of their works; however, it should not sacrifice the important goal of facilitating cross-cultural exchange online. Thus, policymakers should allow free use of AI translation programs, while limiting the amount and type of content that may be translated automatically online in conjunction with software designers building AI translation platforms to reflect these constraints; in other words, programs should respect copyright by adopting a fair use design. Lawmakers should also require AI translation services to allow authors to opt out of having their webpages translated automatically. Third, policymakers should publish norms and guidelines clarifying the extent to which the Fair Use Doctrine covers AI online translation services and when infringements of the law occur. The industry should then update its practices to adhere to the rules. Fourth, given the internet's intrinsically global reach, and given that AI translation programs raise copyright concerns in many jurisdictions, the World Intellectual Property Organization (WIPO) should develop guidelines for these programs' use. United States lawmakers should closely examine the ways in which various jurisdictions have addressed conflicts between new technology online and copyright law and consider how to involve the international community in a workable solution. This framework will close the current gap in the enforcement of copyright law and better preserve the balanced interests the law has long sought to achieve.

These proposals are based on the well-established need to align the general public's right to access knowledge and the arts with the rights of authors to profit from their original creative endeavors for a limited time.[19] Careful calibration of these rights is crucial. Insufficient protection of authors discourages them from publishing their works, thus creating a loss

---

19. *Id.*

for global society.[20] However, unlimited copyright protection of original works may stymie technological progress.[21]

Finally, this paper challenges an underlying premise popular among those who argue wholesale against any enforcement of copyright limitations on automated translation programs: namely, that linguistic variation is an unnecessary encumbrance that should be vanquished whenever and however possible. While translations play an important role in communicating and sharing knowledge, culture, and experience, the success and value of translations hinge on a cultural sensitivity that machines still lack.[22] We argue that making translations widely accessible without paying attention to the quality of the language that actually conveys the information within the underlying text not only infringes on authors' rights, but also ignores the beautiful complexity of human language—far from being a problem, such complexity merits celebration.

In summary, the first Part of this Article traces the controversial history of translation and discusses the elements of good translations. The second Part examines the mechanics of AI translation systems and their ability to learn as they accumulate data, which users increasingly contribute, as well as the limitations of AI translation. The third Part addresses the ways in which AI translations may be biased. The fourth Part explores how the Derivative Works Doctrine has been applied to translations done by humans and considers the implications of classifying AI translations as derivative works. The fifth Part re-examines copyright law and its theoretical justifications in both the United States and other jurisdictions and challenges lawmakers to better regulate the use of AI translation systems to protect the interests of various stakeholders. The final Part posits the need for new legal and digital tools to address the copyright problems posed by AI translation online.

---

20. Glynn S. Lunney, Jr., *Reexamining Copyright's Incentives-Access Paradigm*, 49 VAND. L. REV. 483, 485 (1996) ("[D]efining the appropriate scope of copyright has entailed an examination of incentives and access. Broadening the scope of copyright increases the incentive to produce works of authorship and results in a greater variety of such works. Broadening copyright's scope, however, also limits access to such works both generally, by increasing their price, and specifically, by limiting the material that others can use to create additional works. Given these competing considerations, defining copyright's proper scope has become a matter of balancing the benefits of broader protection, in the form of increased incentive to produce such works, against its costs, in the form of lost access to such works.").

21. *Id.*

22. *See infra* notes 67–71.

II. TRANSLATION IN PERSPECTIVE: FROM THE BIBLE TO GOOGLE
TRANSLATE

Translation has a long and contentious history. As translator David Bellos remarked, "War, diplomacy, trade, and exploration are activities where trust is both crucial and difficult to grant . . . [i]f you don't know the language of your enemy or your partner, you depend entirely on people who do—and there's nothing like dependency to foster resentment and fear."[23] The perception of translation as a tool to build bridges between cultures has never been universally adopted. Taking a historical view allows us to consider just how heated fights over the legitimacy of translation have been.

A. The Controversial History of Translation

The Bible is among the most read, sold, and translated books in the history of the world.[24] It is also among the most controversial, with suspicion running rampant among believers who fear that the translations and interpretations of opposing sects aim to undermine the true faith.[25] Yet, "much of the Western theory and practice of translation stems immediately from the need to disseminate the Gospels."[26]

As the printing press made it easier to share written knowledge in the fifteenth and sixteenth centuries, tensions arose between European Catholics and members of the Protestant Reformation.[27] William Tyndale, the first man to produce a complete English version of the Bible in print, was strangled and burned at the stake in 1536 for a translation that challenged the Roman Catholic Church.[28] The same year, Jacob van Liesveldt, who translated the first Dutch Bible, was beheaded.[29] Centuries later, during the Philadelphia Bible Riots, violence erupted between

---

23. DAVID BELLOS, IS THAT A FISH IN YOUR EAR?, at 117 (2011).

24. *Best-Selling Book of Non-Fiction*, GUINNESS WORLD RECS., http://www.guinness worldrecords.com/world-records/best-selling-book-of-non-fiction [https://perma.cc/HJ5L-R56S] ("Although it is impossible to obtain exact figures, there is little doubt that the Bible is the world's best-selling and most widely distributed book.").

25. *See, e.g.*, HARRY FREEDMAN, THE MURDEROUS HISTORY OF BIBLE TRANSLATIONS: POWER, CONFLICT, AND THE QUEST FOR MEANING 14 (1st ed. 2016) ("Although both the Jewish and Christian sources affirmed the miraculous nature of the translation, they did so for very different reasons. Irenaeus had said that it was so the Jewish translators couldn't collude and falsify the Bible. The Jewish sources claimed it was to eliminate the possibility of people reading ideas into the Hebrew text that weren't there.").

26. GEORGE STEINER, AFTER BABEL: ASPECTS OF LANGUAGE & TRANSLATION 257 (3d ed. 1998).

27. *See* Stanislav Getmanenko, *Freedom from the Press: Why the Federal Propaganda Prohibition Act of 2005 Is A Good Idea*, 114 PENN ST. L. REV. 251, 258–59 n.38 (2009) ("The printing press was the 'catalyst for Protestant Reformation, the Renaissance, and the Industrial Revolution.'").

28. FREEDMAN, *supra* note 25, at 1, 112.

29. *Id.*

"nativist" Protestants and the local Catholic community following a bishop's request that Catholic children be allowed to read the Douay-Reims translation of the Bible at school.[30]

Oral translation (i.e., interpretation) has also long played an important role in world affairs.[31] Between the fifteenth and early twentieth centuries, the Ottoman Empire relied on young men trained in Western languages, such as Italian—many of them from Greek-speaking, Roman Catholic communities—to assist in primarily oral, multilingual diplomacy.[32] When these young men needed to prepare written translations of matters at hand, they followed their usual practices for oral interpretation, altering the Turkish pasha's language to a form best suited to achieving his diplomatic goals; the stakes were high, as disloyalty was punishable by death.[33] In 1821, in fact, a Grand Dragoman named Stavraki Aristarchi was hanged for treason after Greek provinces of the Ottoman Empire revolted: the pasha, it seems, could not trust the translators.[34]

In his memoir *Translating History*, Igor Korchilov recounted his years spent as a Russian-English conference interpreter who worked extensively with the United Nations and Soviet leaders, such as Mikhail Gorbachev, during the Cold War.[35] According to Korchilov, a fundamental quality for interpreters is "what in Russian is called *perevoploscheniye*, or the ability to put oneself in the speaker's shoes, to become his döppelganger, as it were, to catch the essence of his message and not simply repeat what he is saying in another language."[36] Yet, interpreters also face immense pressure when their interlocutors are negotiating nuclear peace and national security; Korchilov describes one Russian interpreter who, while interpreting a speech by Ronald Reagan for Gorbachev and other Russian leaders in 1987 in Washington, D.C., translated the word *adversaries* in the phrase, "a coming together not of allies, but of adversaries," as "competitors."[37] Not unlike the Turkish interpreters from centuries past, this interpreter was "[o]bviously aware of the potential fallout the direct translation of the word would have had on the Soviet audience" at a time when all were seeking to end the Cold War.[38]

---

30. *See generally* Amanda Beyer-Purvis, *The Philadelphia Bible Riots of 1844: Contest Over the Rights of Citizens*, 83 PA. HIST. 366 (2016) (discussing the causes and context of the Philadelphia Bible Riots).

31. *See* BELLOS, *supra* note 23, at 121.

32. *Id.* at 122–25.

33. *Id.*

34. *Id.* at 127.

35. IGOR KORCHILOV, TRANSLATING HISTORY (1997).

36. *Id.* at 21.

37. *Id.* at 65.

38. *Id.*

Translation disputes persist today. In 2017, *The New York Times Magazine* published a lengthy article about a new English translation of the *Odyssey* by classicist Emily Wilson, in which she explained her controversial choice of the word "complicated" for the ancient Greek *polytropos* in the story's famed opening lines, where other translators had used more elaborate, if no less uniform, phrases such as "for wisdom's various arts renown'd," "of wide-ranging spirit," and "so wary and wise."[39] The political nature of translations also received attention when United States President Donald Trump gave a speech before the United Nations General Assembly, and an Iranian interpreter took significant liberties in his rendition of the remarks in Farsi.[40]

The Wycliffe Global Alliance, which aims to make the Bible accessible to as many people as possible in as many languages as required, notes that translation:

> is more than finding out what the biblical text says and saying it in the other language. Translation is a process of discovering what was understood in the original setting and discerning the degree to which readers or hearers in the new context will be able to arrive at a similar understanding.[41]

Translating, therefore, is not—and never has been—a neutral activity; how a concept is converted from one language to another shapes reader or listener perceptions and may serve political ends.

B. Language Barriers and Copyright

The Babel story's sense of awe and frustration with language barriers does not stand alone. In the late nineteenth century, well before the existence of online AI translation (or virtually any automated translation), a Polish physician named Ludwig L. Zamenhof introduced Esperanto to

---

39. Wyatt Mason, *The First Woman to Translate the 'Odyssey' Into English*, N.Y. TIMES MAG. (Nov. 2, 2017), https://nyti.ms/2iTUP84 [https://perma.cc/7D7Y-5P2E]. The journalist who interviewed Wilson observed, "You might be inclined to suppose that, over the course of nearly half a millennium, we must have reached a consensus on the English equivalent for an old Greek word, *polytropos*. But to consult Wilson's 60 some predecessors, living and dead, is to find that consensus has been hard to come by." *Id.*

40. Chris Bell et al., *Iranian Interpreter Defends Trump Speech Omissions*, BBC (Sept. 21, 2017), http://www.bbc.com/news/world-middle-east-41347217 [https://perma.cc/Q79B-2B8X]. In this case, "[The Islamic Republic of Iran] has turned a wealthy country, with a rich history and culture, into an economically depleted rogue state whose chief exports are violence, bloodshed and chaos" was translated to "[i]n our opinion, the life of Iranians could be better." *Id.*

41. *Bible Translation Statistics – Frequently Asked Questions*, WYCLIFFE GLOBAL ALLIANCE (Nov. 2013), http://resources.wycliffe.net/statistics/2013/WycliffeGA_stats_FAQs_EN.pdf [https://perma.cc/7QHC-T5CJ].

the world.[42] Zamenhof hoped that Esperanto, an artificial language that drew heavily from Romance languages, would promote greater intercultural understanding.[43] Though the name "Esperanto" is derived from the Latin root for "hope," the language never flourished.[44] Today, Google Translate has taken up the torch by encouraging users to "[s]peak with the world."[45] Meanwhile, Microsoft Translator recently included Tamil in its repertoire in an effort to "help[] break the language barrier of communication worldwide and within the Indian subcontinent itself."[46]

Writing about the copyright issues of online translation systems over a decade ago, Erik Ketzan argued that these systems "increase social interaction across cultures and languages, possibly dramatically."[47] Therefore, he maintained that such programs should be legally non-infringing, or at least facilitated by the creation of effective licenses, noting that "[i]t is easy to imagine a future in which we seamlessly surf the entirety of the internet in our native language, with occasional (or frequent) linguistic mistakes being the only indication that we are reading machine-translated text."[48] We agree that AI translation systems have great potential. They can facilitate international communication in some settings and perhaps limit the human tendency to demonize others.[49] Accessible AI translation systems can also help people forge economically beneficial partnerships and collaborations. From a social perspective, accessible translation systems can be highly valuable tools, especially in the present globalized era, in which many people work for international

---

42. Olivia B. Waxman, *The Serious History Behind Esperanto*, TIME (July 26, 2016), http://time.com/4417809/esperanto-history-invention [https://perma.cc/4GTZ-KU2W] ("[A] wave of anti-Semitism underscored Zamenhof's thinking that the world needed a single language that would make it possible for people to bridge gaps of religion or ethnicity.").

43. *Id.*

44. *Id.*

45. *About*, GOOGLE TRANSLATE, https://translate.google.com/intl/en/about [https://perma.cc/8NGE-KZJG].

46. Microsoft Translator, *Microsoft Translator Adds Tamil as a Supported Language*, MICROSOFT: MICROSOFT TRANSLATOR BLOG (Oct. 24, 2017), https://blogs.msdn.microsoft.com/translation/2017/10/24/microsoft-translator-adds-tamil [https://perma.cc/XA25-PU7R].

47. Erik Ketzan, *Rebuilding Babel: Copyright and the Future of Online Machine Translation*, 9 TUL. J. TECH. & INTELL. PROP. 205, 218 (2007).

48. *Id.*

49. TORGEIR M. HILLESTAD, U. STAVANGER, THE METAPSYCHOLOGY OF EVIL: MAIN THEORETICAL PERSPECTIVES, CAUSES, CONSEQUENCES AND CRITIQUE 43 (2014), https://brage.bibsys.no/xmlui/bitstream/handle/11250/274771/Rapport_47.pdf?sequence=1 [https://perma.cc/B3SW-44M8]. "In fact, a main function of the group itself is often to build up a barrier between the 'good-us' and the 'bad-others' . . . . We systematically strip the 'others', the opposition, of any human qualities for which we might have empathy or compassion. The 'others' are thus demonized as the very embodiment of evil." *Id.*

companies outside their home country.[50] Our wide-scale multilingual sharing of information is unprecedented and has no doubt exposed people to material they might otherwise never have encountered.[51]

However, unlike Ketzan, we believe that AI translation systems' complete eschewing of copyright law is just one of their significant, though infrequently discussed, drawbacks. Rather than blindly favoring technology over all other rights, we argue that the intellectual property framework requires a new model to address the competing interests at stake. More than a decade after Ketzan published his piece, there is still a huge legal vacuum in this field. Users routinely violate the intellectual property rights of third parties when they engage with AI online translation systems to transfer, for example, full webpages from one language to another, while also freely giving away their data to for-profit corporations in the process, with no real understanding of how companies use that data and how long it is stored. Privacy concerns aside, the continued inaccuracies of AI translation are also problematic, both for authors—whose works the process of translation may deform—and for users, who may unwittingly make offensive statements stemming from algorithmic bias. Human language and intercultural communication, it seems, do not easily fit into the tidy compartments of AI translation.

Additionally, as Ketzan acknowledges, by reducing the world's wealth of languages to a single, off-key tongue, AI translation systems may discourage users from experiencing "foreign travel, music, and literature in the original language"[52] to the detriment of collective diversity and inner enrichment, as culture is strongly intertwined with language.[53] Essentially, although linguistic differences may frustrate us, they also foster innovation and allow societies around the world to find unique cultural expression. In our enthusiastic embrace of new technologies, we should not lose sight of the value of difference, even when it poses challenges.

---

50. *See, e.g.*, Micheline Maynard, *A Lifeline Not Made in the U.S.A.*, N.Y. TIMES (Oct. 17, 2009), https://www.nytimes.com/2009/10/18/business/18excerpt.html [https://perma.cc/U7EB-8KFU] (discussing the "millions of Americans" who work for foreign companies operating in the United States).

51. *See* Margot E. Kaminski & Dr. Shlomit Yanisky-Ravid, *The Marrakesh Treaty for Visually Impaired Persons: Why A Treaty Was Preferable to Soft Law*, 75 U. PITT. L. REV. 255 (2014) [hereinafter Kaminski & Yanisky-Ravid, *Marrakesh Treaty*].

52. Ketzan, *supra* note 47, at 218–19.

53. "One of the best ways to learn objective facts about a culture is to study its language, as language courses tend to immerse students in multiple aspects of a society's objective culture, including its history, literature, and art. Knowledge of a language also provides insight into a social group's perception of reality." Walter A. Wright, *Practical Steps for Acquiring Cross-Cultural Negotiation Skills*, 70 TEX. B.J. 590, 590 (2007).

C. What Makes Translation Good? When is AI Translation Bad?

Italian writer Italo Calvino—who translated the works of French author Raymond Queneau[54]—once observed that "translating is the real way to read a text . . . for an author, reflecting on the translation of a text he has written, discussing it with the translator, is the real way to read his own work, to understand what he has written and why."[55] In his book, *Dire Quasi la Stessa Cosa: Esperienze di Traduzione* ("*Saying Almost the Same Thing: Experiences in Translation*"), Umberto Eco mused that if we can agree that a good translation says "almost the same thing" as its source, we are still left pondering "how elastic must that 'almost' be?"[56] The sensitivity to the language and style required for translation is evidenced by the number of translators who were also prolific writers, including Calvino, Eco, Charles Baudelaire, Cesare Pavese, Vladimir Nabokov, Jose Luis Borges, and Dorothy L. Sayers.[57]

Good translations allow people to share ideas, creativity, criticism, culture, and literature with a wider audience, improving access to knowledge and debate. The world would be a darker place if English speakers could not read *Anna Karenina* and Spanish scientists could not study the works of Albert Einstein. Quality and accuracy matter, however, and largely depend on the fluency of translators in the *cultures* of the languages they are working with.[58] When people communicate, they do so by making certain statements and leaving others unsaid; what a reader understands or a listener hears will often depend on community norms.[59]

---

54. RAYMOND QUENEAU, I FIORI BLU (Italo Calvino trans., 2014).

55. ITALO CALVINO, MONDO SCRITTO E MONDO NON SCRITTO 80 (2002).

56. UMBERTO ECO, DIRE QUASI LA STESSA COSA: ESPERIENZE DI TRADUZIONE 10 (Bompiani 4th ed. 2016).

57. *See About Dorothy L Sayers*, DOROTHY L. SAYERS SOC'Y, https://www.sayers .org.uk/biography [https://perma.cc/P83G-SYQC]; Richard D.E. Burton, *Charles Baudelaire*, ENCYCLOPÆDIA BRITANNICA, https://www.britannica.com/biography/Charles-Baudelaire [https:// perma.cc/D6FT-3VCZ]; *Cesare Pavese*, ENCYCLOPÆDIA BRITANNICA (Sept. 5, 2019), https://www.britannica.com/biography/Cesare-Pavese [https://perma.cc/23MD-9KLY]; Huw Nesbitt, *Jorge Luis Borges's Lost Translations*, GUARDIAN: BOOKS BLOG (Feb. 19, 2010), https://www.theguardian.com/books/booksblog/2010/feb/19/jorge-luis-borges-di-giovanni [https:// perma.cc/8QHV-C3J3]; Vladimir Nabokov, *The Art of Translation*, NEW REPUBLIC (Aug. 4, 1941), https://newrepublic.com/article/62610/the-art-translation [https://perma.cc/5XBE-RPLC] .

58. *See* Sarah Yates, *Scaling the Tower of Babel Fish: An Analysis of the Machine Translation of Legal Information*, 98 L. LIBR. J. 481, 482 (2006) ("Anyone who has attempted to translate a document by simply looking up each word in a bilingual dictionary realizes that this method does not work. The translator must understand the language at the morphological, lexical, and syntactical levels; some would argue that understanding at the cultural level is equally indispensable.").

59. *See* Gideon Toury, *The Nature and Role of Norms in Translation*, *in* THE TRANSLATION STUDIES READER 198 (Lawrence Venuti ed., 2000) ("Translation activities should rather be regarded as having cultural significance. Consequently, 'translatorship' amounts first and foremost to being able to *play a social role*, i.e., to fulfil a function allotted by a community . . . in a way which is deemed

Japanese culture, for example, considers saying "no" directly to be rude; yet the country's cultural norms still afford citizens a linguistic means to decline dinner invitations, for example, in a way that would-be hosts may congenially comprehend.[60] Good translators are not only sensitive to cultural differences, they are also fluent enough in both languages to capture variations in formality, humor, idiomatic expressions, the positive or negative connotations of words, and the musicality of certain sounds, among other important factors.[61] They make deliberate choices about whether and to what extent a text should retain foreign sounds, images, and words, and if a text should read as though it had not been translated at all.[62]

In addition, languages are constantly evolving, and our understanding of words continues to change over time.[63] For example, in the early twentieth century in the United States, the word "Negro" was considered polite and unremarkable—it was, after all, in the name of the United Negro College Fund, and the title of eminent black historian and civil rights activist W.E.B. Du Bois's study, *The Philadelphia Negro*—but today it is perceived as dated, condescending, and racist.[64] As linguist George Steiner noted, "[a]ny thorough reading of a text out of the past of one's own language and literature is a manifold act of interpretation."[65] Like readers of major works of literature, good translators must pay attention to the time in which a text was written in order to capture an author's intended meaning.[66] All of these tasks are difficult, if not

appropriate in its own terms of reference."). *See generally* LIVES IN TRANSLATION: BILINGUAL WRITERS ON IDENTITY AND CREATIVITY (Isabelle de Courtivron ed., 2003).

60. JON P. ALSTON & ISAO TAKEI, JAPANESE BUSINESS CULTURE AND PRACTICES: A GUIDE TO TWENTY-FIRST CENTURY JAPANESE BUSINESS 43 (2005) ("It is best not to ask a Japanese a question demanding either a 'yes' or 'no' answer. Because Japanese do not like to say 'no,' such a question is answered either by a 'yes' or by a vague statement that does not obligate the speaker. Saying 'no' bluntly is considered impolite and aggressive.").

61. *See, e.g.*, APRILIA ZANK, THE WORD IN THE WORD: LITERARY TEXT RECEPTION AND LINGUISTIC RELATIVITY 158 (2013) ("The first prerequisite for a competent translator is, doubtlessly, very good command of both the target language . . . and the source.").

62. *See* BELLOS, *supra* note 23, at 44–59.

63. Robin I. M. Dunbar, *The Origin and Subsequent Evolution of Language*, *in* LANGUAGE EVOLUTION (Morten H. Christiansen and Simon Kirby eds., 2003) ("Language has two remarkable properties. First, it allows us to communicate ideas with each other; second, languages evolve and diversify with a speed and facility that is quite unique within biological evolution.").

64. *See, e.g.*, Tess Owen, *The Words 'Oriental' and 'Negro' Can No Longer be Used in US Federal Laws*, VICE (May 21, 2016), https://news.vice.com/article/the-words-oriental-and-negro-can-no-longer-be-used-in-us-federal-laws [https://perma.cc/3F6R-XCQH] ("The law targets two anti-discrimination areas of US law that use antiquated language to describe racial or ethnic groups.").

65. STEINER, *supra* note 26, at 18. This work provides a lengthy analysis of the significance of languages and translation.

66. *See* Peter Messent, *Censoring Mark Twain's 'N-Words' Is Unacceptable*, GUARDIAN (Jan. 5, 2011), https://www.theguardian.com/books/booksblog/2011/jan/05/censoring-mark-twain-n-word-unacceptable [https://perma.cc/K79H-3AW8] (arguing that although Mark Twain used the word

impossible, for an AI translation program to achieve with fluency because such work requires in-depth understanding, not merely a statistical analysis of data.

In fact, although AI translation tools have dramatically improved and continue to evolve, they are often unreliable and might always remain so. Yehoshua Bar-Hillel, a well-known Israeli mathematician and linguist, spent years working at the Massachusetts Institute of Technology (MIT) developing machine translation, but even he expressed doubts about the feasibility of fully automatic, high quality translation systems.[67] "By not realizing the practical futility" of the goal of perfect machine translation, Bar-Hillel argued that its enthusiasts had misled themselves, and the agencies sponsoring their research, into believing that the "real thing" (the perfect and efficient translation) was just around the corner.[68] Bar-Hillel stressed that it would have made more sense to consider the potential applications for imperfect automated translation systems.[69]

Although nearly sixty years have passed since Bar-Hillel published his paper in 1960, not much has changed. Recently, translator Douglas Hofstadter illustrated the inability of Google Translate to accurately translate simple sentences that would not have posed much of a challenge to sufficiently fluent human translators.[70] The problem, he argued, was that "[m]achine translation has never focused on understanding language. Instead, the field has always tried to 'decode'—to get away without worrying about what understanding and meaning are."[71] A lack of precision, an inability to perceive nuance, and the lack of an "ear" for language might afflict some human translators, but these problems are inherent for machines. Understanding the fascinating processes underlying AI translation systems, which we discuss later in this Article, helps explain the persistence of these problems.[72]

The standards of reliability used for the translation of court documents provide a good metric for evaluating the reliability of AI translation systems from a legal perspective. If the latter is appropriate only in limited circumstances, where quality and accuracy are of

---

"nigger" many times in *Huckleberry Finn*, the novel should be understood as anti-racist and the word should not be edited out).

67. Yehoshua Bar-Hillel, *A Demonstration of the Nonfeasibility of Fully Automatic High Quality Translation*, in LANGUAGE AND INFORMATION 174–79 (1964). *See generally* Yehoshua Bar-Hillel, *Some Linguistic Problems Connected with Machine Translation*, 20 PHIL. SCI. 217 (1953).

68. Bar-Hillel, *supra* note 67.

69. Bar-Hillel, *supra* note 67.

70. Douglas Hofstadter, *The Shallowness of Google Translate*, THE ATLANTIC (Jan. 30, 2018), https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570 [https://perma.cc/Y2X9-HX4J].

71. *Id.*

72. *See infra* Part III.

secondary importance, then we ask: can we appropriately implement this policy guideline to make such distinctions regarding the web, where AI systems now routinely provide coarse translations of copyrighted material, to the general public? In this sense, we argue that the pervasive inaccuracies of AI translation should inform any discussion about balancing access to knowledge with the rights of authors in the copyright regime. If AI translation cannot be consistently relied upon for formulaic texts, then when it is used for more colloquial and colorful works, the underlying copyright violation becomes even more pertinent because a poor quality translation may violate an author's moral rights in places where these are legally recognized.[73] Yet web users rely on AI translation for all sorts of material, and this issue gets scant attention.

Another related problem relates to the biases inherently embedded in AI systems in general.[74] This issue is crucial when we discuss AI translation systems' lack of linguistic understanding. Professionals in a wide range of industries have greeted the arrival of AI systems enthusiastically, focusing on their efficiency and ability to process vast troves of data.[75] AI systems are now involved in hiring and firing, granting loans, evaluating tenants, and deciding which airline passengers should be on the No Fly List.[76] Yet, despite their usefulness, AI systems can also

---

73. "The non-economic interests of authors are found worthy of protection because of the presumed intimate bond between authors and their works, which are almost universally understood to be an extension of the author's personhood. The standard set of moral rights recognized in the literature consists of the author's right to claim authorship (right of attribution), the right to object to modifications of the work (right of integrity), the right to decide when and how the work in question will be published (right of disclosure), and the right to withdraw a work after publication (right of withdrawal)." Cyrill P. Rigamonti, *Deconstructing Moral Rights*, 47 HARV. INT'L L.J. 353, 355–56 (2006).

74. Shlomit Yanisky-Ravid & Sean Hallisey, *"Equality and Privacy by Design": A New Model of Artificial Intelligence Data Transparency via Auditing, Certification, and Safe Harbor Regimes*, 46 FORDHAM URB. L.J. 428 (2019) (describing many types of biases in AI systems).

75. "Generally, big data analytics focuses on examining large data sets to gather useful information to help organizations make more-informed business decisions. Businesses are capitalizing on this newfound ability by overhauling the ways in which they source and evaluate candidates, analyze attrition risks, optimize individual and team performance, identify safety risks, provide wellness guidance and resources, and much more." Nathaniel M. Glasser et al., *Big Data, People Analytics and Employment Decisions: The Rewards and Often-Overlooked Risks*, WESTLAW J. EMP., Mar. 27, 2018, at 1.

76. *See* Yanisky-Ravid & Hallisey, *supra* note 74 (noting that data can be illegal, discriminatory, manufactured, unreliable, or simply incomplete, and that the more data AI systems "swallow," the greater the likelihood they will contribute to biased, discriminatory decisions or could violate privacy. The article discusses how discrimination can arise, even inadvertently, from the operation of "trusted" and "objective" AI systems, and advocates for increased transparency); *see also* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1297 (2008); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 636 (2017) ("The tools currently available to policymakers, legislators, and courts were developed to oversee human decisionmakers and often fail when applied to computers instead. For example, how do you judge the intent of a piece of software?

generate biased results that violate fundamental rights or, at the very least, flout social norms.[77] Safiya Umoja Noble, for example, has highlighted the sexist and racist failures of algorithmic data.[78] Privacy scholar Danielle Keats Citron has argued that automated systems' lack of notice and transparency jeopardizes due process norms and further, that "[a]utomation encourages agencies to adopt overly simplified policy, which can more easily be translated into code."[79] In the translation context, while Turkish uses the same pronoun, "o" for he and she, until recently, when Google Translate was asked to translate "o bir doctor" (he or she is a doctor) into English, it defaulted to the masculine "he is a doctor."[80]

We do not pretend that human translators are free of biases; any choice of words reflects personal perspective and experience. However, human translators do understand the meaning of the words before them and are capable of making nuanced distinctions. The now pervasive presence of AI translation services, their reliance on opaque data, and the influence of automation bias—that is, the global populace's general trust of algorithmic results—all suggest that our enthusiasm for these products should not blind us to their weaknesses.[81]

Because automated decision systems can return potentially incorrect, unjustified, or unfair results, additional approaches are needed to make such systems accountable and governable.").

77. *See* Kroll et al., *supra* note 76, at 636; *see also* Lee Rainie & Janna Anderson, *Code Dependent: Pros and Cons of the Algorithm Age*, PEW RES. CTR. (Feb. 8, 2017), http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age [https://perma.cc/4JVK-NY4Z].

78. SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM 4 (2018) (arguing that search algorithms produce racist results because they reflect the biases of the people who develop them).

79. *See* Citron, *supra* note 76, at 1297.

80. *See* Li Zhou, *Is Your Software Racist?*, POLITICO (Feb. 7, 2018), https://www.politico.com/agenda/story/2018/02/07/algorithmic-bias-software-recommendations-000631 [https://perma.cc/B93E-JK4K].

> Google's Translate tool "learns" language from an existing corpus of writing, and the writing often includes cultural patterns regarding how men and women are described. Because the model is trained on data that already has biases of its own, the results that it spits out serve only to further replicate and even amplify them. It might seem strange that a seemingly objective piece of software would yield gender-biased results, but the problem is an increasing concern in the technology world . . . Voice-based assistants, like Amazon's Alexa, have struggled to recognize different accents. A Microsoft chatbot on Twitter started spewing racist posts after learning from other users on the platform. In a particularly embarrassing example in 2015, a black computer programmer found that Google's photo-recognition tool labeled him and a friend as "gorillas."

Id. For further discussion of this issue, see James Zou & Londa Schiebinger, *AI Can Be Sexist and Racist—It's Time to Make it Fair*, 559 NATURE INT'L J. SCI. 324 (July 12, 2018).

81. *See* Citron, *supra* note 76, at 1271–72 ("[H]uman beings rely on automated decisions even when they suspect system malfunction. The impulse to follow a computer's recommendation flows from human 'automation bias'—the 'use of automation as a heuristic replacement for vigilant information seeking and processing.' Automation bias effectively turns a computer program's suggested answer into a trusted final decision.").

E. Translation and Copyright Protection

The right of authors to original and derivative copyrights in their works is guaranteed internationally. For example, the Berne International Convention for the Protection of Literary and Artistic Works (Berne Convention) defines translation as a protected derivative work that authors own and control.[82] The United States Copyright Act of 1976 also recognizes translation as a derivative work, in which the author holds exclusive rights.[83] Copyright's legal protections even extend to intangible goods presented to the public, and there is no exception for the internet: copyright law protects works even when they are accessible to all in cyberspace, and even when they can be easily downloaded and copied.[84] No registration or special display symbol is necessary; once fixed in "any tangible medium of expression,"[85] authors have automatic copyright protection for their original creative works.[86] In other words, the mere fact that copyrighted works exist online in an easy-to-access format does not mean they are not protected or that they have become public domain.

When authors' works appear on the web, they serve as a constant source of free material for large, for-profit entities such as Google and Microsoft, who make significant financial gain from their translation—precisely the type of exploitation that copyright law seeks to prevent. Because the internet is global, by examining the theoretical justifications for copyright law and the relevant legal norms in four countries—the United States, the United Kingdom, Canada, and France—this Article brings a wider comparative and global view of copyright protection in the 3A era. First, however, we address the evolution of AI systems in general and of AI translation systems specifically.

---

82. "Authors of literary and artistic works protected by this Convention shall enjoy the exclusive right of making and of authorizing the translation of their works throughout the term of protection of their rights in the original works." Berne Convention for the Protection of Literary and Artistic Works art. 8, Sept. 9, 1886, S. Treaty Doc. No. 99-27 (1986) [hereinafter Berne Convention].

83. 17 U.S.C. § 101 (2018).

> A "derivative work" is a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a "derivative work."

Id.; *see also* 17 U.S.C. § 103 (2018).

84. *See generally* Robin Andrews, Note, *Copyright Infringement and the Internet: An Economic Analysis of Crime*, 11 B.U. J. Sci. & Tech. L. 256, 256 (2005).

85. *See* 17 U.S.C. § 102 (2018).

86. "[R]egistration is not a condition of copyright protection." 17 U.S.C. § 408 (2018).

III. The Development of AI Translation Systems

Though definitions of AI vary, AI systems are generally understood as human-like systems that can perform complex tasks involving decision-making, creativity, learning, development, and communication—systems that usually require human intelligence.[87] Such systems are presently used in a variety of fields, from diagnostic medicine and automated weaponry to intelligence gathering, and from the world of art and inventions to that of contract drafting.[88]

A. Artificial Intelligence (AI) Systems: What, When, How

As computers grew in popularity throughout the 1940s and became commercially available, their brain-like attributes caught the popular imagination.[89] In 1946, the press described the Electronic Numerical Integrator and Computer (ENIAC) as a "giant brain," and, thereafter, computers regularly outperformed humans in mathematical exercises and games such as chess.

In his well-known article *Computing Machinery and Intelligence*, computer scientist and mathematician Alan Turing considered the ability of machines to think, observing that "[i]ntelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation, but a rather slight one, which does not give rise to random behaviour, or to pointless repetitive loops."[90] Nearly seventy years later, technological advances have brought AI programs increasingly into the public awareness, yet even experts in the field struggle to agree on a precise definition for AI, largely because intelligence is an ambiguous concept even for humans.[91] Today, instead of emphasizing machines' abilities to perform tasks, AI definitions tend to "focus on the concept of machines that work to achieve goals."[92]

---

87. *See* Person, *supra* note 7; *see also* Yanisky-Ravid & Liu, *supra* note 7.

88. *See* Shlomit Yanisky-Ravid & Luis Antonio Velez-Hernandez, *Copyrightability of Artworks Produced by Creative Robots and Originality: The Formality-Objective Model*, 19 Minn. J.L. Sci. & Tech. 1, 4–6 (2018).

89. Ray Kurzweil, How to Create a Mind: the Secret of Human Thought Revealed 180 (2013) ("When computers first became a popular topic in the 1940s, they were immediately regarded as thinking machines. The ENIAC, which was announced in 1946, was described in the press as a 'giant brain.' As computers became commercially available in the following decade, ads routinely referred to them as brains capable of feats that ordinary biological brains could not match.").

90. A. M. Turing, *Computing Machinery and Intelligence*, 49 Mind 433, 459 (1950).

91. *See* Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J.L. & Tech. 353, 359 (2016) ("The difficulty in defining artificial intelligence lies not in the concept of artificiality but rather in the conceptual ambiguity of intelligence.").

92. *Id.* at 361.

AI systems can be described as those "capable of performing tasks that would normally require human intelligence, [evidenced by] recognition, decision-making, creation, learning, evolving, and communicating," or that allow humans to reach their goals more efficiently by processing data.[93] While early AI "expert systems" ran by applying rules input by humans, today AI systems function via "neural networks" that imitate human brain functions by analyzing patterns within the data they process and learning through trial and error.[94] For example, in 2003, Raymond Kurzweil patented his Cybernetic Poet, a program which analyzes poems written by human authors, creates a word-sequence model, and generates new, original verse in the same style.[95] The Next Rembrandt project sought to expose a computer to as many digital copies of Rembrandt's works as possible in order to teach it not only to replicate those paintings, but also to generate new ones with a similar look without any further input from the software developers and art historians.[96] Famed auction house Christie's sold its first piece of AI-generated portraiture for $432,500.[97] Nowadays, AI systems also produce music and many other sorts of creative works.[98]

---

93. Yanisky-Ravid, *Generating Rembrandt*, *supra* note 7, at 673.

94. *Id.* at 675.

95. Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5, 36.

> Raymond Kurzweil was granted United States Patent No. 6,647,395 for [his] . . . Cybernetic Poet, [which] is designed to be used either as a "poet's assistant" or as an automatic poetry generator. The program is "provided with an input file of poems written by a human author or authors. It analyzes these poems and creates a word-sequence model based on the poems it has just read. It then writes original stanzas of poetry using the model it has created." The resulting stanzas, according to the patent's written description, "will have a similar style to the poem(s) originally analyzed and contained in the author analysis model, but will be *original poetry generated by the process*."

Id.

96. *See* Yanisky-Ravid, *Generating Rembrandt*, *supra* note 7, at 663–64 (describing the AI-produced Rembrandt Project and the questions of ownership and accountability it raises).

97. The work sold for nearly forty-five times its high estimate. *See Is Artificial Intelligence Set to Become Art's Next Medium?*, CHRISTIE'S (Dec. 12, 2018), https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx [https://perma.cc/TZ4J-LB7X] (noting that the AI system produced the work after algorithmic exposure to 15,000 others); *see also* Yanisky-Ravid*, Generating Rembrandt*, *supra* note 7.

98. For a demonstration of jazz music produced by an AI system, see *AI and Intellectual Property Law Featuring Prof. Shlomit Yanisky-Ravid*, FORDHAM INTELL. PROP. MEDIA & ENT. L.J. PODCAST (May 4, 2018), http://www.fordhamiplj.org/2018/05/04/episode-51-ai-and-intellectual-property-law [https://perma.cc/DYW2-XSEY] (featuring audio of a jazz composition generated by an AI system at the 14:05 mark); *see also, e.g.*, Dani Deahl, *How AI-Generated Music is Changing the Way Hits are Made*, VERGE (Aug. 31, 2018), https://www.theverge.com/2018/8/31/17777008/artificial-intelligence-taryn-southern-amper-music [https://perma.cc/PK9D-HXC7] (noting that "music-making AI software has advanced so far in the past few years that it's no longer a frightening novelty; it's a viable tool that can and *is* being used by producers to help in the creative process").

The publishing industry is also experimenting with AI uses in journalism; a *Los Angeles Times* journalist developed an algorithm called Quakebot, which produces simplistic news stories based on United States Geological Survey reports, and companies such as Narrative Science and Automated Insights have also experimented with AI uses for the news business.[99] Though AI news stories still remain on the fringes of the news market and do not always sound quite right to readers, interest in the potential applications for the evolving technology still remains high.[100]

Despite the overwhelming enthusiasm for AI systems, some commentators have raised concerns about the ways in which they produce biased results and may perpetuate undesirable forms of discrimination by relying on unrepresentative data sets.[101]

B. AI Translation Systems

AI translation systems are computer systems which can translate content with or without human assistance; they are not translation tools, such as online dictionaries or terminology forums.[102] The systems may be designed to be bilingual or multilingual, unidirectional or bidirectional.[103] The oldest type of translation system functions with a "direct translation" approach: designed for a specific pairing of languages, it operates by moving from the source language unidirectionally to a target language

---

99. JAQUES BUGHIN ET AL., 2017 MCKINSEY GLOB. INST., ARTIFICIAL INTELLIGENCE: THE NEXT DIGITAL FRONTIER?, at iv (2017), https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx [https://perma.cc/7PD6-G62D].

> Artificial intelligence is poised to unleash the next wave of digital disruption, and companies should prepare for it now . . . . AI investment is growing fast, dominated by digital giants such as Google and Baidu. Globally, we estimate tech giants spent $20 billion to $30 billion on AI in 2016, with 90 percent of this spent on R&D and deployment . . . early AI adopters that combine strong digital capability with proactive strategies have higher profit margins and expect the performance gap with other firms to widen in the future . . . . Early adopters are already creating competitive advantages, and the gap with the laggards looks set to grow.

Id.; *see also* Robert C. Denicola, *Ex Machina: Copyright Protection for Computer-Generated Works*, 69 RUTGERS U. L. REV. 251, 257 (2016) ("Whatever the future holds, artificial intelligence is already an active participant in the creative process.").

100. Denicola, *supra* note 99, at 261. *See generally* Lin Weeks, Note, *Media Law and Copyright Implications of Automated Journalism*, 4 N.Y.U. J. INTELL. PROP. & ENT. L. 67 (2014) (describing uses of AI in journalism).

101. *See generally* Michele Willson, *Algorithms (and the) Everyday*, 20 INFO., COMM. & SOC'Y 137 (2016).

102. *See generally* W. John Hutchins, *Machine Translation: A Brief History*, *in* CONCISE HISTORY OF THE LANGUAGE SCIENCES: FROM THE SUMERIANS TO THE COGNITIVISTS 431–45 (Ernst F. K. Koerner & Ronald E. Asher eds., 1995).

103. *Id.*

using certain rules.[104] An example of a programming rule might specify that the English "you" can be either "vous" or "tu" in French.[105] A second type of translation system, known as "interlingua," relies on two stages of translation: first, from a source language into an intermediate "interlingua," and then, from the interlingua into the target language.[106] This type of system may rely on an artificial, or auxiliary, language as a cushion between the source and target languages.[107] It digitally identifies patterns and similarities among all languages and then creates a new language that synthesizes all common structures before translating the data to the target language. A third type of translation system follows a transfer approach: converting source language texts into abstract representations, then converting into the target language equivalent representations, and finally, into the target language text.[108] In each of these methods, the AI system relies on a constant input of new data, meaning that the data provider or machine "trainer" are a vital part of the system's functioning.

The first automated translation systems began to appear in the 1930s and gradually improved throughout the remaining decades of the twentieth century, particularly in the United States, Soviet Union, Japan, and Europe.[109] Commercial interest in translation gained momentum throughout the 1990s alongside the rise of the internet.[110]

Today, AI translation systems are everywhere: on mobile phones, computers, virtual assistants, and many other devices.[111] Business use of online, simultaneous translation continues to expand as the available systems improve. Users, in fact, play an active role in the development of AI translation tools by providing feedback and language corrections.[112] The following discussion will focus on this unique role of users around the globe, and their important contribution to the development of AI translation systems, which evolve with every additional piece of data.

C. Users as Data and Feedback Providers

A unique and interesting feature of AI online translation systems is the involvement and contribution of the users in improving and programming them. AI systems function thanks to an array of stakeholders, which one of this Article's authors, Dr. Yanisky-Ravid, has

---

104. *Id.*
105. Ketzan, *supra* note 47, at 212.
106. Hutchins, *supra* note 102.
107. *Id.*
108. *Id.*
109. *Id.*; *see also* Ketzan, *supra* note 47, at 209.
110. *See* Ketzan, *supra* note 47, at 210.
111. *See* Gershgorn, *supra* note 9.
112. *See infra* Part III.

referred to as "the multi-player model."[113] The players in this model include software developers, data suppliers, trainers who provide feedback, owners and operators, manufacturers and even, possibly, the state and insurance companies.[114] Because these systems rely on large volumes of data, data providers perform a vital role in the development of any AI system.[115] Likewise, AI translation systems depend on receiving a massive amount of data and feedback in all languages, without which they cannot function properly.[116] To meet this need, AI translation is remarkable in its reliance on crowd wisdom.[117] In other words, by using AI translation systems, we all become players, improving them by providing data and feedback. Thus, we not only use and enjoy the systems, we also refine their abilities. Google Translate, for one, has encouraged its users around the world to be data providers and AI trainers, who correct the system and help fine-tune its translations, thus transforming users into an integral part of the multi-player model.[118] The boundaries between users and system developers blur in never-ending, evolving cycles of machine learning and improving. The significant contributions of AI translation program users might, however, have legal consequences from a copyright perspective.[119]

In some respects, Google Translate and similar programs mirror the work of human translators. The machines look at a sentence and must decide on a likely appropriate equivalent in the target language based on data previously encountered from users.[120] Moreover, AI systems' ability to process vast amounts of data at a high volume in some contexts makes them more efficient than human translators.[121] AI translation tools rely on a translation memory, "a database that stores previously translated *segments*," and before they translate a file of segments, they analyze how

---

113. *See* Yanisky-Ravid, *Generating Rembrandt*, *supra* note 7, at 691–93 (discussing questions of ownership and accountability for works produced by AI systems); *see also* Yanisky-Ravid & Liu, *supra* note 7, at 2231–36 (suggesting that the involvement of multiple stakeholders in inventions created by AI systems disrupts the traditional patent process because there is no single inventor).

114. *See* Yanisky-Ravid & Liu, *supra* note 7, at 2232–33.

115. Yanisky-Ravid & Hallisey, *supra* note 72 (focusing on data as the main source of bias in algorithmic results).

116. *Id.*

117. For a discussion of crowd wisdom, see, e.g., Ralph Hertwig, *Tapping Into the Wisdom of the Crowd—With Confidence*, 336 Sci. 303, 303–04 (2012).

118. *See* Google Translate Community, https://translate.google.com/community [https://perma/cc/UW2C-Y7G4].

119. *See infra* Part IV (examining AI translations according to the derivative works doctrine).

120. *See, e.g.*, Andy Way, *A Critique of Statistical Machine Translation*, *in* Linguistica Antverpiensia, New Series-Themes in Translation Studies 17, 25 (2009) ("In a nutshell, the goal of PB-SMT is to find the most likely translation *T* of a source sentence *S*. We say 'most likely,' as many possible candidate target language translations may be proposed by the system.").

121. Yanisky-Ravid, *Generating Rembrandt*, *supra* note 7.

many translated segments are available for reuse.[122] Similarly, a human translator must read a sentence and draw on previous experience and knowledge of the target language to propose a translation that sounds as smooth and similar to the original as possible.[123]

Nevertheless, the question of whether AI online translation systems can ever truly replace the work of human translators remains open. Word meanings are context-dependent, and people who speak several languages have a nuanced cultural understanding of written expression that eludes even the most high-level machines.[124] AI translation systems—unlike human beings—plug in their solutions based on data input, following what might be described as the "Chinese Room" method, meaning that a correct output can result without knowing whether the machine understands the meaning as humans do (or think they do).[125] By contrast, human translators actually understand thoughts, emotions, and the words they read, and can, therefore, better grapple with ambiguities and have a higher potential for accuracy.[126]

In an article comparing accuracy assessments for machine translation with similar assessments for human translators, linguists and researchers Marina Fomicheva and Lucia Specia described a conceptual flaw: machine translations are usually evaluated in comparison to a human translation, instead of to their original source text—which is the standard method for reviewing human translations.[127] Implicit in this approach is the

---

122. Christophe Eyraud, *Case Study: Cutting 15% of Translation Costs with Artificial Intelligence*, MEMSOURCE (May 29, 2018), https://www.memsource.com/blog/2018/05/29/case-study-cutting-15-of-translation-costs-with-artificial-intelligence [https://perma.cc/W3Y5-Z4GZ].

123. *See, e.g.*, Lawrence Venuti, *The Translator's Invisibility*, 28 CRITICISM 179, 179 (1986) (arguing that people commonly feel that "the more 'successful' the translation, the more invisible the translator, and the more visible the author or meaning of the original text").

124. *See* EUGENE A. NIDA, CONTEXTS IN TRANSLATING 11 (2001) ("[A] language is always part of a culture and the meaning of any text refers directly or indirectly to the corresponding culture.").

125. *See* Turing, *supra* note 90, at 460. Turing proposed that a human have separate conversations on a screen with a machine and with another person, and then guess which was which; if the human could not reliably identify the machine, the machine was said to have passed the test. *Id. But cf.* John Searle, *Minds, Brains, and Programs*, 3 BEHAV. & BRAIN SCI. 417 (1980). In this famous paper about the "Chinese Room," Searle argued that machines can massive amounts of data without understanding and without thinking, and therefore the "Turing Test" cannot be used to conclude that a machine can think. *Id.* For a discussion of the Personhood and Consciousness Approach to AI Systems, see Yanisky-Ravid, *Generating Rembrandt*, *supra* note 7, at 684–85 (describing the notion that perception works the same way for humans and AI systems).

126. Nevertheless, human translators can make mistakes, too. *See, e.g.*, KORCHILOV, *supra* note 35, at 20 (describing a mistake by an interpreter at a major international conference, who caused considerable upset between Soviet and Danish delegates after he mistranslated a Russian saying that literally means "an elderberry grows in the kitchen and my uncle lives in Kiev" and figuratively means "mixing apples and oranges," into "something is rotten in the kingdom of Denmark").

127. Marina Fomicheva & Lucia Specia, *Reference Bias in Monolingual Machine Translation Evaluation*, 2 54TH ANNUAL MEETING OF THE ASS'N FOR COMPUTATIONAL LINGUISTICS: PROCEEDINGS OF THE CONFERENCE 77, 77–82 (2016).

assumption "that the task of translation has a single correct solution," when in fact, "except for very short sentences or very specific technical domains, the same source sentence may be correctly translated in many different ways."[128] Thus, even in the way it is assessed, AI translation lags behind the richness and greater accuracy of human translation.

In sum, AI online translation systems have many advantages. Compared to human translators, they are faster, more instantly accessible, cheaper, and convenient in that they operate in almost every language. However, even though human translation has never been perfect, AI online translation systems lack some significant human abilities and therefore present major drawbacks, limitations, and new challenges, all of which will be discuss in the following subsection.

D. The Limitations of AI Translation

Throughout history, translations have been both an important communication tool and a source of tension.[129] In his book, *Through the Language Glass*, Israeli linguist Guy Deutscher muses over the following questions: "[i]s [language] an artifact of culture or a bequest of nature? If we hold language up as a mirror to the mind, what do we see reflected there: human nature or the cultural conventions of our society?"[130] One might ask related questions of translations: how closely do they capture the cultural ideas and imagery expressed in the original works they are based on? Must they be done by humans to properly convey meaning, or are machines just as capable of reflecting nuanced ideas?

Over the past several decades, there has been an increase in the global use of online translation services. AI translation systems go far beyond Google's famous translation tool, encompassing a variety of services. Not only can web users read content translated into multiple languages by pasting text into a box or by asking the service to translate an entire web page, but they can also access lectures, movies, songs, and books—which are just as easily translated into any widely spoken language—in a single click.[131] Many AI web translation services are freely accessible, while some are only available to software brand customers.[132] Yet, because AI translation services rely on data and statistics rather than personal understanding, they cannot replicate all the nuances of human languages, idioms, and colloquial expressions, so AI-translated results are often a

---

128. *Id.* at 77.

129. *See generally* FREEDMAN, *supra* note 25.

130. GUY DEUTSCHER, THROUGH THE LANGUAGE GLASS: WHY THE WORLD LOOKS DIFFERENT IN OTHER LANGUAGES 9 (2010).

131. *See generally* Ketzan, *supra* note 47.

132. *Machine Translation*, MICROSOFT, https://www.microsoft.com/en-us/translator/mt.aspx [https://perma.cc/E7HG-RMSF].

distorted reflection of original works. For example, AI translations of web pages are frequently rough and inaccurate, while AI translations of copyrightable books or papers contain many mistakes.[133] We argue that, given the polemics that have so often surrounded translations throughout history, future copyright disputes over the unauthorized linguistic efforts of machines are inevitable.

AI translation relies on technology to convert text from one language into another by continuously collecting linguistic data and sorting it, creating new and more accurate patterns with each additional piece of data.[134] Although AI translations have greatly improved over the twentieth and twenty-first centuries, significant differences in quality and accuracy between machine and human translations still remain. While these differences might diminish over time, they are unlikely to completely disappear, especially as some languages have been translated more than others, leading to differences in the availability of training data for machines.[135] In short, AI translations remain unsophisticated compared to those produced by people.[136] This inherent drawback raises questions about appropriate uses of translation and how such uses may affect the market for authors', as well as human translators', copyrighted works.

In some contexts, the rough-around-the-edges nature of AI translations poses no great problem, as the stakes are low and, therefore, less harmful. In such contexts, once the threats of AI translation have been acknowledged, these systems can be efficiently used. Reading a street sign or menu, driving with a taxi driver who does not understand your language, or asking for directions in a foreign country, all present examples of how AI translation is poised to efficiently facilitate communication, as compared to alternative, more threatening scenes. The internet is replete with gems such as "Take Luggage of Foreigner, No Charge."[137] If one's goal is merely to broadly comprehend the subject

---

133. *See* Ketzan, *supra* note 47, at 224 (noting that machine translation, "even excellent . . . has little to offer the fully bilingual human translator. The reason is that MT has a 'tin ear,' so to speak").

134. *See* Yates, *supra* note 58, at 484.

135. *See* VR Ferose, Barney Pell & Lorien Pratt, *From the Business of Language to the Language of Business: The Future of Translation Worldwide*, DIGITALIST MAG. (May 17, 2018), https://www.digitalistmag.com/machine-learning-ai/2018/05/17/future-of-translation-worldwide-06168565 [https://perma.cc/8EFK-R9DE] ("These less-translated languages have less training data, making the automation problem harder . . . .").

136. *See* Ian Henderson, *Chasing the Unicorn of Auto-Translation: Why Automation Tools Can't Work for Businesses*, WIRED (Aug. 2014), https://www.wired.com/insights/2014/08/chasing-unicorn-auto-translation-automation-tools-cant-work-businesses [https://perma.cc/D9UW-EJZ3] (noting that "'language' is far more than just words" and that "[f]or machines to accurately translate, there needs to be as much abstract thinking as mathematical calculations").

137. Chungcharn Or, *Lost in Translation: Our Top 20 Picks*, LONELY PLANET (Nov. 24, 2011), https://www.lonelyplanet.com/blog/2011/11/24/lost-in-translation-our-top-20-photos [https://perma.cc/C8ZU-EM9L].

matter of a text, then AI translation works quite well. The European Patent Office (EPO) has even partnered with Google Translate to create an automated translation service for public patent documents.[138] In this instance, the stated goal is to provide people with the ability to skim documents in various languages for a basic understanding of their content.[139] "The machine translation system should give you the gist of any patent or patent-related document, and help you to determine whether it is relevant. You might decide on this basis whether you need to invest in a human translation of the document," the EPO advises, adding: "Please note that the engine cannot provide legally binding translations."[140] In other words, as long as "the gist" of meaning suffices, AI translation can provide a valuable service and facilitate efficient research. In *Found in Translation: How Language Shapes Our Lives and Transforms the World*, authors Nataly Kelly and Jost Zetzsche came to the same conclusion, noting, "[F]or some projects, companies just want an approximate translation—they don't care if the clothes fit properly, so long as they can be worn . . . . Automated translation does have its place, especially in certain industries."[141]

Problems arise, however, when the quality and accuracy of a translation are more important. In the medical field, where doctor-patient communication is vital to the provision of care, errors in automated translation can have serious consequences.[142] United States judges have held that Google-translated documents do not satisfy the proper authentication requirements for extradition under 18 U.S.C. § 3190[143] and have expressed strong skepticism of the accuracy of AI translation.[144] Linguist David Bellos observed that "you should never use [Google

---

138. *Patent Translate*, Eur. Patent Off., http://www.epo.org/searching-for-patents/helpful-resources/patent-translate.html#tab3 [https://perma.cc/J9RZ-XS53].

139. *Id.*

140. *Id.*

141. Nataly Kelly & Jost Zetzsche, Found in Translation: How Language Shapes Our Lives and Transforms the World 219 (2012).

142. For a discussion of the pros and cons of using AI translation systems in the medical field, see Rachel M. Taylor, Nicola Crichton, Beki Moult & Faith Gibson, *A Prospective Observational Study of Machine Translation Software to Overcome the Challenge of Including Ethnic Diversity in Healthcare Research*, 2 Nursing Open 14 (2015).

143. *See, e.g.*, In re Extradition of Ferriolo, 126 F. Supp. 3d 1297, 1301 (M.D. Fla. 2015) (holding that an internet search engine's English translations of Italian criminal statutes were inadmissible in court).

144. Martinez v. United States, 828 F.3d 451, 483 (6th Cir. 2016) (Clay, J., dissenting) ("The majority's argument is less than persuasive when one realizes that its legal argument is predicated on a popularized and less than precise translation source. Language is complex and the meaning of a word can vary greatly depending on its context, as any native speaker can attest. It would seem that the majority is defining various phrases in a vacuum without considering the appropriate contexts in which they were used.").

Translate] to translate into a language you do not know very well. Use it only to translate into a language in which you are sure you can recognize nonsense."[145]

Ultimately, who should be responsible for the results of AI translation? How concerned should we be about the biases these results may contain? Who owns these new works? Who should benefit from them? These questions will be discussed first according to the traditional, derivative works approach, then through an examination of theoretical justifications for intellectual property law and, finally, according to the new model we propose in this study.

E. The Hidden Biases of AI Translation Systems and Their Undesirable Results

Like other AI systems, AI translation systems offer great benefits—for instance, facilitating communication when perfect accuracy is not required—however, they also pose challenges related to their inability to perceive or understand nuance, as well as to the bias in their algorithms.[146] Thus, some of the same risks associated with AI systems in the employment or financial sectors also apply to online translation services. Because large firms dominate the AI sector, AI translation systems' results may reflect the biases of these companies, or of society at large, if users provide feedback, as they can for Google Translate.[147] If training data is flawed, discriminatory outcomes can occur, perpetuating society's discriminatory attitudes.

In an article entitled *Accountable Algorithms*, the authors discuss the lack of accountability mechanisms in place for software that makes decisions that formerly would have been made by individuals, arguing that the law has lagged behind as technology has leapfrogged ahead.[148] In addition to advocating for greater collaboration between lawmakers and computer scientists, the authors highlight some of the problems that can emerge when algorithmic bias goes unchecked, noting that "[a] significant concern about automated decisionmaking is that it may simultaneously systematize and conceal discrimination," and further, that "decisions made by computers may enjoy an undeserved assumption of fairness or

---

145. *See* BELLOS, *supra* note 23, at 256.

146. *See, e.g.*, Willson, *supra* note 101, at 9 ("[T]he outcomes point to the lack of nuance or contextual understanding that such processes often do not accommodate, and the impact they may have when being addressed to complex, socially embedded human users and systems. They are also not open to outside scrutiny that might enable underpinning assumptions to be interrogated: it is only their products or outputs that can be addressed.").

147. *See supra* note 118.

148. *See generally* Kroll et al., *supra* note 76.

objectivity."[149] For example, a program designed to select the best applicants for a job may be trained using information about candidates a company has previously hired; if those candidates were predominantly male—not because they were better-qualified but because the former human resources director was sexist—then the algorithm will simply perpetuate the original bias, favoring male applicants over similarly qualified female applicants.[150] Even when no malice is involved, an algorithm's distorted analysis of statistics can lead to unfair results, for instance, by directing police officers to stop and frisk members of certain groups because, historically, those groups have "been the target of disproportionate enforcement."[151]

Although many scholars see the software as the main source of biases,[152] a recent study suggested that the data, rather than the software, played a much more important role in creating and spreading biases.[153] According to this view, exposing AI systems to more data does not necessarily eliminate biases because society itself is biased, and the data therefore reflects sexism, racism, and a myriad of other social biases. Writing in *Nature*, an international journal of science, James Zou and Londa Schiebinger describe Google Translate converting Spanish news articles into English, transforming "she said" into "he said," and a Nikon camera software perceiving Asians as always blinking.[154] They argue that:

> A major driver of bias in AI is the training data. Most machine-learning tasks are trained on large, annotated data sets . . . [i]n natural-language processing, standard algorithms are trained on corpora consisting of billions of words. Researchers typically construct such data sets by scraping websites, such as Google Images and Google News, using specific query terms, or by aggregating easy-to-access information from sources such as Wikipedia. These data sets are then annotated, often by graduate students or through crowdsourcing platforms such as Amazon Mechanical Turk. Such methods can unintentionally produce data that encode gender, ethnic and cultural biases. Frequently, some groups are over-represented and others are under-represented.[155]

---

149. *Id.* at 680.

150. *Id.* at 681.

151. *Id.*

152. *Id. See generally* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018) (arguing in favor of revealing algorithms in criminal trials, even when they involve trade secrets).

153. Yanisky-Ravid & Hallisey, *supra* note 74 (arguing that data, rather than software, is the main cause for the perpetuation of social biases and rights violations by AI systems).

154. Zou & Schiebinger, *supra* note 80.

155. *Id.*

Because faulty algorithms can magnify biases through feedback loops, Zou and Schiebinger note that each time a program such as Google Translate defaults to a masculine pronoun, it increases that pronoun's relative frequency on the web, "potentially reversing hard-won advances towards equity."[156]

Though linguist Bellos humorously advised Google Translate users to steer clear of the program unless they were familiar enough with their target language to recognize "nonsense,"[157] the same users should also be on the lookout for bias. Many of those consulting AI translation services are, nonetheless, wholly unfamiliar with the languages for which they are seeking assistance and are, consequently, oblivious to any offense they may cause. We argue that when AI online translation systems switch male and female pronouns or use racist expressions, they are accurately mirroring social prejudices; however, instead of simply accepting and parroting back these prejudices, we would like to see the systems taking a more progressive linguistic stance. Previously, Yanisky-Ravid has argued that, when necessary, AI systems should disclose the data they rely on, and that an auditing and certification program for algorithms would incentivize industry to implement non-discriminatory best practices. Google is reportedly working to address its bias problems by offering both masculine and feminine translation options when both are appropriate, but so far, the improvements are limited. They are partially available, for example, only "when translating English into French, Italian, Portuguese or Spanish and when translating from Turkish to English."[158] How effective Google's updates will be across its entire language repertoire remains to be seen.

Regardless, the problem is that when users rely on AI online translation systems to convert a known language into an unknown one, they have no way of assessing the results and any biases remain concealed. These hidden biases are not the only problem with AI online translation systems.

IV. Translation and Copyright Law: Past, Present, and Future

Translations undoubtedly require creativity, but are they best understood as original or derivative works? The legal doctrine surrounding translations has changed over time, highlighting some of the tensions present in the current copyright issues of AI-generated translations. Here, we examine these questions more closely and conclude that if translations

---

156. *Id.*

157. BELLOS, *supra* note 23, at 256.

158. Mallory Locklear, *Google Is Working to Remove Gender Bias in Its Translations*, ENGADGET (Dec. 7, 2018), https://www.engadget.com/2018/12/07/google-remove-gender-bias-translations [https://perma.cc/2UKT-YX7J].

continue to be classified as derivative works, a new legal and digital framework is needed to facilitate the use of AI translation systems while respecting the balancing goals of copyright law.

A. The Past: Translation as a New Creative Work

Authors have not always had the right to control the production of derivative works based on their original creations: in early copyright disputes, courts held that translating books without an author's authorization was *not* infringement.[159] For many years, judges perceived translations, abridgements, and improvements "as intellectual products of the second comers' creativity" and believed that such new works "contributed to the advancement of knowledge—the very goal of modern copyright law."[160]

In the UK, some believed that a translation of a book fell outside the scope of the 1710 Statute of Anne, which provided for court enforcement of copyright law.[161] In a famous English case, *Burnett v. Chetwood*, the Lord Chancellor agreed with the defendants, who stood accused of producing an unauthorized translation of a copyrighted book, that the statute "could be intended only to restrain the mechanical art of printing, and that others should not pirate the copy and gain an advantage to themselves by reprinting it" and should not apply to a translation of the book, "which in some respects may be called a different book, and the translator may be said to be the author, in as much as some skill in language is requisite thereto, and not barely a mechanic art, as in the case of reprinting in the same language."[162] In other words, because translation requires great skill and does not involve rote copying, it is better viewed as original work and thus the translation of a copyrighted work should not require the permission of the original author.

Similarly, in the American case *Stowe v. Thomas*, the author Harriet Beecher Stowe sued a man for translating her copyrighted work, *Uncle Tom's Cabin*, into German and publishing it in chapters in a Philadelphia newspaper, *Die Freie Presse*.[163] She had already hired and compensated a different translator, and the authorized translation was published and copyrighted.[164] The court held that the unauthorized translation did not

---

159. *See* Pamela Samuelson, *The Quest for a Sound Conception of Copyright's Derivative Work Right*, 101 GEO. L.J. 1505 (2013) (referencing Stowe v. Thomas, 23 F. Cas. 201, 208 (C.C.E.D. Pa. 1853) (No. 13,514) and Burnett v. Chetwood, 35 Eng. Rep. 1008 (Ch. 1720)).

160. Samuelson, *supra* note 159, at 1507.

161. Copyright Act of 1710 (Statute of Anne), 8 Ann. c. 21 (Gr. Brit.); *see also Burnett*, 35 Eng. Rep. 1008.

162. *Burnett*, 35 Eng. Rep. 1008.

163. *Stowe*, 23 F. Cas. 201.

164. *Id.*

violate Stowe's rights as an author because it was not a copy, but rather an inventive work that borrowed her ideas, which were not subject to copyright:

> [t]o make a good translation of a work, often requires more learning, talent and judgment, than was required to write the original. Many can transfer from one language to another, but few can translate. To call the translations of an author's ideas and conceptions into another language, a copy of his book, would be an abuse of terms, and arbitrary judicial legislation.[165]

Nevertheless, in 1870 the United States Congress amended copyright law, providing authors with more control over the translation of their works, and in 1909, it established a careful list of derivatives that authors were allowed to manage.[166] There was a sense that unauthorized translations unjustly enriched the infringer and hurt demand for the original work, even if unauthorized translations were most likely "to supplant demand for authorized derivatives than to supplant demand for the original."[167] Since a proliferation of bootleg translations would undermine authors' discretion in choosing how to render their original works in new markets, the law was changed to grant authors greater control over translations of their works.[168]

In *Shook v. Rankin*, an 1875 case from the Northern District of Illinois, complainants were the owners of an original, unpublished French manuscript that had been translated into English by a man named Jackson with the consent of the original authors.[169] Defendants were performers who had acted out Jackson's translation at a Chicago theater without his permission.[170] They denied that Jackson had acquired any rights by his translation or copyright and argued that they intended to use a different translation, also produced with the consent of the original authors.[171] The court enjoined defendants from using Jackson's translation, but did not extend the injunction to the other translation.[172] Noting that as an authorized translator of the play, Jackson was "in the sense of the law, the

---

165. *Id.* at 207.

166. *See* Samuelson, *supra* note 159, at 1508–09.

167. *Id.* at 1519.

168. *See* Daniel Gervais, *The Derivative Right, or Why Copyright Law Protects Foxes Better Than Hedgehogs*, 15 VAND. J. ENT. & TECH. L. 785, 799 (2013) ("When copyright moved doctrinally beyond literal copying, the underlying concern was the (correct) realization that other appropriations, including translation, dramatizations, and substantial (if not literal) copying, could cause an unjustified harm to a copyright holder.").

169. Shook v. Rankin, 21 F. Cas. 1335, 1336 (C.C.N.D. Ill. 1875).

170. *Id.*

171. *Id.*

172. *Id.*

author of that for which he obtained a copyright," Circuit Judge Drummond said that he failed to see why Jackson "was not protected under the law for his translation and adaptation of the work to the stage, and *of which he was in one sense the author*."[173]

Debates over the appropriateness of giving authors control over translations of their works have not been fully extinguished. Today, with authors' permission, translators can obtain the copyright to their translations, which then have the status of original, rather than derivative works.[174] In a 1987 dispute between the writer of a biography of Igor Stravinsky and the composer's former personal assistant, the latter asserted his inherited copyright to the English translations from Russian of some of Stravinsky's correspondence.[175] Though the plaintiff spoke no Russian, he had hired some Russian-speaking students to do initial, literal translations of the correspondence, and "thereafter he reworked the language, exercising his own authorship to achieve a diction he considered suitable," recasting the translations "in such a way as to imitate the unusual manner in which Stravinsky spoke English."[176] The court agreed that Stravinsky's former assistant had a strong copyright claim to the translations, either under a theory that his reworking of them was substantial and original or under a "work for hire" theory involving the students.[177]

Translations of religious works have long been, and continue to be, a source of conflict. Agreeing that a translation of an English prayer-book into Hebrew was sufficiently original to be copyrightable, the court in *Merkos L'Inyonei Chinuch, Inc. v. Otsar Sifrei Lubavitch, Inc.* said that "[t]he translation process requires exercise of careful literary and scholarly judgment."[178] Further, in *Society of Holy Transfiguration Monastery, Inc. v. Gregory*, a Greek Orthodox monastery that had produced original translations of ancient religious works brought a copyright infringement suit against an archbishop who had left the order and posted some of the translations on his own website.[179] One of the archbishop's arguments, in his defense, was that the works at issue were not original and therefore not copyrightable—that they were "merely a commixture of various ancient religious works that had previously been translated into English" with

---

173. *Id.* (emphasis added).

174. 17 U.S.C. §§ 101, 103 (2018).

175. Craft v. Kobler, 667 F. Supp. 120, 121–22 (S.D.N.Y. 1987).

176. *Id.* at 125.

177. *Id.*

178. Merkos L'Inyonei Chinuch, Inc. v. Otsar Sifrei Lubavitch, Inc., 312 F.3d 94, 97 (2d Cir. 2002).

179. Soc'y of Holy Transfiguration Monastery, Inc. v. Gregory, 689 F.3d 29, 35 (1st Cir. 2012).

minor changes.[180] The Monastery argued that its translations were both original and entitled to copyright protection under the Copyright Act's "express acknowledgement of the copyrightability of derivative works."[181] The court agreed that translations may be original and copyrightable despite being derivative, as long as "there is originality in [the works'] arrangement or selection."[182] The court also rejected the archbishop's argument that the "merger doctrine," by which a copyright may be held invalid if there is only a limited number of ways to express a given idea, as "inapposite" because "the art of translation involves choices among many possible means of expressing ideas."[183]

By contrast, in *Signo Trading Intern. Ltd. v. Gordon*, a translator who had been commissioned by the manufacturer of an electronic translation system to translate English words into Arabic did not have a copyright claim because the lists of words did not have the requisite degree of originality and because the words on the list had been chosen by the manufacturer.[184]

Echoing some of the views underlying the *Chetwood* and *Stowe* opinions, in *The Scandals of Translation: Towards an Ethics of Difference*, translator, translation theorist, and historian Lawrence Venuti argues that translations are a unique category of derivative work in that they do not involve literally copying a foreign text, but instead enter "into a mimetic relation that inevitably deviates from the foreign language by relying on target-language approximations."[185] Far from limiting the market for an author's work, Venuti suggests that translations may potentially expand it "by demonstrating its value abroad."[186] He advocates modifying the current copyright regime to incentivize translators, so that if an author has not made use of translation rights within five years of publishing a work, the first translator or publisher to translate that work should be entitled to enjoy an exclusive translation right in the work for as long as the translation remains in print.[187]

We argue that debates over the originality or derivative nature of translations have thus far considered only human translators, but an urgent need still remains to consider how applicable notions of originality and derivation are to AI translations, for reasons we will discuss further.

---

180. *Id.* at 46.

181. *Id.*

182. *Id.* at 47.

183. *Id.* at 53.

184. Signo Trading Int'l Ltd. v. Gordon, 535 F. Supp. 362, 363 (N.D. Cal. 1981).

185. LAWRENCE VENUTI, THE SCANDALS OF TRANSLATION: TOWARD AN ETHICS OF DIFFERENCE 64 (1998).

186. *Id.*

187. *Id.* at 65–66.

B. The Present: Translation as a Derivative Work and the Author as Owner

Currently, both United States law and international intellectual property law view translations as derivative works, that is, as new products containing significant portions of existing, copyrightable works.[188] In other words, a translated work is considered the original work dressed in a different cloak.[189] Whereas in the past, translations were necessarily the work of multilingual individuals, in today's 3A era, translations may also be the product of popular AI translation programs, especially online.[190] However, machine translations continue to lack the sophistication of those produced by humans.[191]

There are five legal implications of defining translated works as derivative. First, users of AI translation systems become infringers, subject to legal penalties, when they translate copyrighted works without authors' permissions. Second, anyone who wishes to translate a copyrighted work must seek permission from the work's author, who may grant a license for the translation of the work.[192] Third, users of AI translation may have to pay authors for these licenses. Fourth, the author owns the resulting translation, which is the legal default even when an author has granted permission for a work to be translated; of course, the parties can negotiate another outcome. Fifth, as long as the author owns the rights to a translation, the author is entitled to any benefits resulting from the translation; thus, the user should pay royalties to the author.[193]

In the United States, copyright protection extends to all "original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or

---

188. *See, e.g.*, Jason L. Cohn, *The King James Copyright: A Look at the Originality of Derivative Translations of the King James Version of the Bible*, 12 J. INTELL. PROP. L. 513, 516 (2005) ("The main distinction between derivative works and original works is that original works are almost entirely created from materials and ideas that originated with the author while derivative works are based primarily on materials and ideas derived from other sources.").

189. *See* Copyright Act, 17 U.S.C. § 102 (2018); Berne Convention, *supra* note 82.

190. *See, e.g.*, Gideon Lewis-Kraus, *The Great A.I. Awakening*, N.Y. TIMES MAG. (Dec. 14, 2016), https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html [https://perma.cc/Q97E-RFKA].

191. *See* Henderson, *supra* note 136 (noting that "'language' is far more than just words" and "[f]or machines to accurately translate, there needs to be as much abstract thinking as mathematical calculations").

192. *See Business Considerations in Translations*, *in* 2 LINDEY ON ENTERTAINMENT, PUBLISHING AND THE ARTS § 5:140 (3d ed. 2018) ("[T]ranslators are authors who create new works but whose rights are normally subordinated to the rights of the author of the underlying work. Most frequently, translators are engaged under work for hire agreements and obtain no copyright interest in their works.").

193. *Id.*

device."[194] Little is required to satisfy the notion of originality in United States copyright law, as the Supreme Court explained in its cornerstone copyright case, *Feist Publications, Inc. v. Rural Telephone Service Co., Inc.*: "Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity."[195] Examples of original works include literary, musical, dramatic, pictorial, and architectural creations.[196] Internationally, the Berne Convention states that it covers "every production in the literary, scientific, and artistic domain, whatever may be the mode or form of its expression," and cites "books, pamphlets and other writings" as examples.[197] While the Convention does not list originality as a prerequisite for copyright protection, "most member states have an originality and independent creation requirement."[198]

In contrast to an original work, a derivative work is one "based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted."[199] International copyright law considers translation an important form of derivative work.[200] The 1886 Berne Convention first addressed the issue of copyright and translation at the global level, granting authors the exclusive right to make or authorize translations of their works for ten years.[201] The 1896 revision to the Convention extended this right to cover the entire copyright term, and the 1908 revision "finally assimilated the translation right into general reproduction rights."[202]

Among the exclusive rights granted to authors of original works, the Berne Convention lists "the right to translate"[203] and "moral rights" which encompass **"the right to claim authorship of the work and the right to object to any mutilation, deformation or other modification of, or other derogatory action in relation to, the work that would be prejudicial to the

---

194. 17 U.S.C. § 102(a) (2018).

195. Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991).

196. *Id.*

197. Berne Convention, *supra* note 82, at art. 1.

198. Susanna Frederick Fischer, *Dick Whittington and Creativity: From Trade to Folklore, From Folklore to Trade*, 12 TEX. WESLEYAN L. REV. 5, 23 (2005).

199. 17 U.S.C. § 101 (2018).

200. Berne Convention, *supra* note 82, at art. 2.

201. Ketzan, *supra* note 47, at 208–09.

202. *Id.* at 209.

203. Berne Convention, *supra* note 82, at art. 8.

author's honor or reputation."[204] Since translations are derivative, authors may protect the moral rights to their works by giving or withholding their consent to commercial translations of their works.[205]

C. The Future Is Already Here: Are the Products of AI Translation Derivative Works?

Authors own their copyrightable works. Therefore, applying the derivative work doctrine to AI translation means that users of online AI translation programs—including, but not limited to, Google Translate—infringe stakeholders' rights by creating unauthorized derivative works.

We argue that in the 3A era this conclusion faces challenges. First, as globalization has made the notion of community international, online translation tools have become a primary means of participating in cultural life, enjoying the arts, and sharing in scientific advancement, as well as seeking, receiving, and imparting "information and ideas of all kinds."[206] Allowing citizens to become acquainted with poems, stories, news and other written material online through internet access is especially important where access to such material offline may be limited or nonexistent. Following this logic, the international community agreed in the Marrakesh Treaty to create an exception to copyright law so that visually impaired persons, who often lack access to writings, could access copyrighted written materials by using digital tools to transfer the words to audible form.[207] On the other hand, the Universal Declaration of Human Rights also guarantees the moral rights of authors, and even where works are translated by machines, copyright law and the derivative works doctrine are applicable.[208]

AI online translation services are not human; however, humans are the investors in and software designers of AI translation programs, and they turn a profit from the use of these programs in the marketplace. In *Naruto et al. v. David Slater*, the Ninth Circuit, citing the United States Copyright Office, held that non-human animals could not be copyright holders, associating authorship with a person's control over a work.[209] In

---

204. *Summary of the Berne Convention for the Protection of Literary and Artistic Works (1886)*, WORLD INTELLECTUAL PROP. ORG., http://www.wipo.int/treaties/en/ip/berne/summary_berne.html [https://perma.cc/3HW6-7D23].

205. *Id.*

206. *See* Universal Declaration of Human Rights, G.A. Red. 217A (III), U.N. Doc A/810 art. 27, at 71 (1948); *see also* International Covenant on Civil and Political Rights art. 19(2), Dec. 19, 1966, 999 U.N.T.S. 171.

207. *See* Kaminski & Yanisky-Ravid, *Marrakesh Treaty*, *supra* note 51.

208. *See* sources cited *supra* note 125.

209. "We must determine whether a monkey may sue humans, corporations, and companies for damages and injunctive relief arising from claims of copyright infringement . . . [W]e conclude that

Naruto, an Indonesian macaque was found to have no copyright in a selfie he had taken with a photographer's camera because the Copyright Act makes no mention of non-human animals;[210] United States courts have repeatedly referred to "persons" or "human beings" when analyzing copyright authorship. Following the logic of the *Naruto* case, which awarded authorship of the photograph to the photographer who set up the camera, AI online translation authorship lies with the private companies financing and developing the AI software. Naruto was not a participant in the human economy, but companies such as Google are.

Another consideration is the active role users play in generating new AI translations by providing corrections and suggested improvements. There is a strong argument that existing copyright law fails to capture this scenario, which sees many people contributing to a machine's translation. Even if the software that allows a machine to translate is the result of programming and data-feeding by humans, who could be considered the authors of whatever the machine generates, the presence of other contributors muddies the waters. Consider the parallels with open-source software: "Open source licenses . . . allow software to be freely used, modified, and shared."[211] The code in programs such as Linux "is free and available to the public to view, edit, and—for users with the appropriate skills—to contribute to."[212] On AI online translation platforms, should contributors be entitled to a share of the profits from their work? The precise terms under which contributors correct online AI translations are often unclear, yet the companies producing the software systems benefit greatly from user contributions. Another relevant parallel is with the comment sections of major news websites, where individuals contribute their insights regarding current events and drive traffic to the websites. Should they have no intellectual property rights to their comments? We argue that the law is unclear on these issues, which are central to AI online translation platforms.

Finally, the broad, far-reaching use of AI online translation systems on computers and smartphones around the globe makes their post-hoc regulation through the application of existing law difficult, as consumers

---

this monkey—and all animals, since they are not human—lacks statutory standing under the Copyright Act." Naruto v. Slater, 888 F.3d 418, 420 (9th Cir. 2018).

210. *Id.*

211. *See, e.g.*, *Licenses & Standards*, Opensource.org, https://opensource.org/licenses [https://perma.cc/B8GP-X2BS].

212. *See What is Linux?*, Opensource.org, https://opensource.com/resources/linux [https://perma.cc/ST9W-PCBM].

have grown accustomed to using these tools.[213] Enforcing existing copyright law would be akin to banning all music downloads in the Napster case[214]—possible, but what would be the result and would it best serve the balancing goals of copyright law?

Taken together, these challenges posed by AI translation online reveal a glaring legal void. The popularity of AI online translation and its importance should prompt us to consider how the law can efficiently address these issues. There are three approaches to address this question: (1) existing law can sufficiently address these new situations and nothing need be done save enforce copyright law and shut down online AI translation programs; (2) IP laws in general, including copyright laws, are so outdated as to be unusable when examining AI systems;[215] and (3) certain sections of copyright law have become outdated, and amendments to existing law that take technological innovation into consideration are necessary. Consider, for instance, that AI systems such as New Rembrandt, are capable of producing their own art work and inventions; these developments are not unlike AI translation. Yanisky-Ravid and Luis–Antonio Velez–Hernandez previously suggested revising the legal understanding of the term "creativity" so that it might encompass artworks generated by AI systems.[216] Here, we argue that while the balancing goals of intellectual property laws remain relevant, AI systems present such significant challenges to existing legislation and regulation that new legal and digital tools are needed.[217]

V. RETHINKING COPYRIGHT'S APPLICABILITY TO AI TRANSLATION

Current copyright law defines users of AI translation tools as lawbreakers, subject to severe penalties for their infringement. The following discussion re-examines copyright law in the 3A Era and challenges the system to better address and regulate the near-universal use of AI translation systems. We first analyze the use of this advanced technology from the perspective of copyright law before discussing the troubling ramifications of this analysis and how to address them. Like many other web-based products and services, AI translation systems pose

---

213. *See* Alex Suskind, *15 Years After Napster: How the Music Service Changed the Industry*, DAILY BEAST (June 6, 2014), https://www.thedailybeast.com/15-years-after-napster-how-the-music-service-changed-the-industry [https://perma.cc/7K82-LF8E].

214. For further discussion of Napster, see *infra* notes 257–269.

215. *See, e.g.*, Yanisky-Ravid, *Generating Rembrandt*, *supra* note 7, at 663–64.

216. *See* authorities cited *supra* note 88.

217. *See, e.g.*, Shlomit Yanisky-Ravid & Kenneth S. Kwan, *3D Printing, the Road Ahead: The Digitization of Products When Public Safety Meets Intellectual Property Rights—A New Model*, 38 CARDOZO L. REV. 921 (2017) (discussing the intersection of safety and intellectual property concerns related to AI-generated products).

specific copyright conundrums that, left unexamined, lead to a poor balancing of the interests at the heart of intellectual property law.

A. Theoretical Justifications: The Tension Between Authors' Rights and Sharing Knowledge

Justifications for intellectual property laws and copyright take various forms, but all wrestle with the fundamental tension between the rights of creators and those of society at large.[218] This friction has shaped international agreements such as the Berne Convention and the domestic copyright laws of different countries. In this Article, we examine the theoretical underpinnings of intellectual property law and how they apply to AI-generated translation systems. We argue that this analysis provides a useful frame of reference for examining the copyright issues of autonomous AI translation online.

Three traditional, internationally-dominant frameworks exist for justifying or limiting intellectual property laws: the Utilitarian (or Law and Economics) theory; the Labor theory; and the Personality theory. Analysis of these traditional justifications, which all favor giving authors rights in their creative works, indicates that copyright law should protect authors from unauthorized online translation of these works.[219]

However, two other, more recent justifications, lead to different conclusions. The first, the Distributive Justice Theory, stems from John Rawls's theory of justice and challenges the three main frameworks as promoting insufficiently egalitarian outcomes.[220] The second comes from Lior Zemer's work, *The Idea of Authorship in Copyright*, in which he argues in favor of a more collective, or environmental, theory of copyright based on the notion that individual creative processes are shaped by surrounding culture, and therefore, the public at large should be a joint author for every copyrighted work.[221] While we will briefly discuss these approaches to copyright law, it is important to note that they do not reflect existing legislation.

The United States Constitution offers a prime example of the Utilitarian justification for intellectual property rights: Article I, Section 8 grants Congress the power "[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the

---

218. *See, e.g.*, *supra* note 18.

219. *See* Shlomit Yanisky-Ravid, *The Hidden Though Flourishing Justification of Intellectual Property Laws: Distributive Justice, National Versus International Approaches*, 21 LEWIS & CLARK L. REV. 1, 5 (2017) (describing the theoretical justifications for intellectual property and focusing on the distributive justice model).

220. *Id.* at 15–16.

221. *See generally* LIOR ZEMER, THE IDEA OF AUTHORSHIP IN COPYRIGHT (2007).

exclusive Right to their respective Writings and Discoveries."[222] The idea is that the government creates intellectual property rights to provide creative individuals with a strong incentive to share the fruits of their labor with the public at large. "[T]he Framers intended copyright itself to be the engine of free expression. By establishing a marketable right to the use of one's expression, copyright supplies the economic incentive to create and disseminate ideas."[223] Absent such protections, free-riders, who had not invested any efforts, time, money, or creative work could pilfer the ideas of hard-working creative individuals who might be discouraged enough to renounce their imaginative pursuits at a loss to the general public.[224] By granting artists, scientists, or writers exclusive rights over their works for a limited number of years, this approach balances providing creative individuals with benefits and furthering the public good. Currently, copyright in the United States typically lasts for the life of the author plus seventy years, with some variation in the case of works made for hire or those made anonymously, or pseudonymous.[225] According to a Utilitarian theory of intellectual property, any regulation of AI online translation programs must function as an incentive to authors to share their works with the general public.

The Labor theory of intellectual property, which has strongly influenced U.K. law, has Lockean underpinnings. In the Second Treatise on Government, John Locke argued that

> Though the earth and all inferior creatures be common to all men, yet every man has a "property" in his own "person." This nobody has any right to but himself. The "labour" of his body and the "work" of his hands, we may say, are properly his. Whatsoever, then, he removes out of the state that Nature hath provided and left it in, he hath mixed his labour with it, and joined it to something that is his own, and thereby makes it his property . . . For this "labour" being the unquestionable property of the labourer, no man but he can have a right to what that is once joined to, at least where there is enough, and as good left in common for others.[226]

---

222. U.S. CONST. art. I, § 8, cl. 8.

223. Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 558 (1985).

224. *See, e.g.*, Matt Jackson, *Harmony or Discord? The Pressure Toward Conformity in International Copyright*, 43 IDEA 607, 614–15 (2003). "Intellectual property is particularly susceptible to freeriders since a disproportionate share of the expense involved in creating intellectual property goes toward the creation of the first copy. These first-copy costs can only be recouped if the creator can control the distribution of successive copies." *Id.*

225. *How Long Does Copyright Protection Last?*, U.S. COPYRIGHT OFF., https://www. copyright.gov/help/faq/faq-duration.html [https://perma.cc/KS7P-DNWA].

226. JOHN LOCKE, SECOND TREATISE OF GOVERNMENT 17 (2004).

Applied to copyright, this philosophy seeks to balance benefiting authors for their strenuous efforts and ensuring that the world at large is not left empty-handed in the process. In the case of automated online translation, this model suggests that authors' consent is needed because authors have a property right in their labor.

The Personality theory focuses more on the close connection between creators and their works. Under this framework, creators should own the property rights to their creations because these are a form of highly personal self-expression.[227] Civil law countries such as France, Germany, and Italy favor the notion of inalienable moral rights, recognizing that works contain an expression of the personality of their authors.[228] Viewed from this perspective, authors' temporary monopoly on the use of their works stems not from the effort they expended or from a government incentive to produce more, but from the intimate, unique connection authors have with their works. It also means that "corporate entities and employers who hire third parties to create works for them do not qualify as authors" because they were not involved in the deeply individual process of creation.[229] Following this approach, authors should be able to control if and how their works are translated using AI systems because the works are extensions of the authors as individuals.

The Distributive Justice theory challenges the first three by positing that the law should set guidelines to ensure equitable distribution of societal resources.[230] Relying on John Rawls' philosophical "veil of ignorance,"[231] this framework invites policymakers to devise intellectual property laws while imagining that they do not know what social class they belong to and what abilities they have.[232] The idea is that the veil of ignorance guides policymakers toward more naturally fair results.[233]

---

227. *See, e.g.*, Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287, 345 (1988) (describing a property owner's view of "property as an expression of his self").

228. Rigamonti, *supra* note 73, at 359 ("The copyright statutes currently in force in France, Germany, and Italy contain provisions dedicated to the protection of the rights of disclosure, attribution, integrity, and withdrawal. These rights are generally conceptualized as inalienable rights of authors in their works . . . . [M]oral rights are rights of *authors*, which is to say that only those human beings who *actually create* the work in question qualify as owners of moral rights.").

229. *Id.*

230. *See* Yanisky-Ravid, *supra* note 217, at 10.

231. *See* Chris Chambers Goodman, *Beneath the Veil: Corollaries on Diversity and Critical Mass Scholarships from Rawls' Original Position on Justice*, 13 WASH. & LEE J. CIV. RTS. & SOC. JUST. 285, 287 (2007) ("John Rawls identifies the 'original position,' in which all actors are behind a 'veil of ignorance,' such that no one knows what qualities, attributes, privileges, and abilities each person might hold. This veil of ignorance as to differences provides the first level playing field, a vacuum even, from which all can discuss and debate issues of fairness and justice without any influence of self-interest or self-preservation.").

232. *See* Yanisky-Ravid, *supra* note 217.

233. *Id.*

Viewed this way, AI online translation programs should be limited by copyright law only to the extent that this does not cause gross inequalities of access to translated materials.

Finally, the Collective model proposed by Lior Zemer runs counter to the notion that individual authors produce anything truly original in the sense that they are always inspired by what surrounds them.[234] In *The Idea of Authorship in Copyright*, he wrote that "[e]very copyrighted entity represents the creative collectivity. It is a joint enterprise of both the individual author and the public."[235] According to this analysis, copyright law in its existing form is based on a flawed premise of originality, and the translation by AI systems of copyrighted material online should pose no authorship obstacles, particularly as the AI systems are routinely fed on contributions from a large pool of users.

Current law espouses neither the Distributive Justice theory nor the Collective model; however, both frameworks are relevant to the ways in which existing copyright law is ill-equipped to deal with AI translations online.

B. The Fair Use Doctrine and AI Translation Online

Just as the invention of the printing press both facilitated the unauthorized copying of original works and improved public access to written materials, leading to the birth of copyright law, in the twentieth and twenty-first centuries a proliferation of new internet services has led to disputes stemming from the public's unprecedented, widespread access to copyrighted material.[236] The internet has allowed people to share vast quantities of material with one another around the world as never before, facilitating global communication.[237] However, the rapid diffusion of content on the internet has also led to lawsuits as the tension between authors' rights and the public domain takes on new forms.[238] Time and

---

234. LIOR ZEMER, THE IDEA OF AUTHORSHIP IN COPYRIGHT 1 (2007) ("Every copyrighted entity represents the creative collectivity.").

235. *Id.* at 1.

236. *See* Robert L. Shaver, *Copyright Law in the Digital Age*, 49 ADVOCATE 17 (2006) ("Before the printing press, an author was not concerned about the unauthorized copying of original works, because copying was so impractical as to be impossible. In fact, without the printing press, there was no group of authors who made a living by writing. With the printing press it became possible for an author's works to be widely disseminated, and to actually make some money by being an author. With the ability to copy came the need to control who could copy, and copyright law was born.").

237. Am. Civil Liberties Union v. Reno, 521 U.S. 844, 117 (1997) ("It is no exaggeration to conclude that the internet has achieved, and continues to achieve, the most participatory marketplace of mass speech that this country—and indeed the world—has yet seen.").

238. *See* Christopher Gagne, *Canon AEDE: Publishers' Protections From Digital Reproductions of Works by Search Engines Under European Copyright Law*, 29 TEMP. INT'L & COMP. L.J. 203, 203 (2015) ("Google has become a household name throughout the world, offering a virtual 'one stop shop' for answers to questions on every topic imaginable. This convenience, however, comes

again, courts are asked to evaluate whether use of this material is allowed under the fair use doctrine, or whether it undermines the Founders' goal of promoting "the progress of Science and useful Arts."[239]

Fair use is an affirmative defense to claims of direct copyright infringement; the United States Copyright Act of 1976 provides guidance on the American "fair use" doctrine, codified in section 107.[240] It cites four relevant factors for examining whether a particular use of an original work is infringing or not: the purpose and character of the use (commercial or noncommercial, transformative or reproductive); the nature of the copyrighted work; the amount and substantiality of the original work used; and the effect of such use upon the market, or potential market, of the original work.[241] The preamble to the Act's fair use section specifically cites "criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship or research" as valid fair use purposes.[242] Courts conducting a fair use analysis seek to determine "whether the copyright law's goal of promoting the Progress of Science and useful Arts would be better served by allowing the use than by preventing it."[243] A fair use defense "presupposes good faith and fair dealing, and one pertinent consideration is whether the user stands to profit from exploitation of the copyrighted material without paying the customary price."[244] American courts have often viewed commercial uses of copyrighted material as presumptively unfair exploitations of authors' monopoly privileges, while considering noncommercial uses of such material, especially in the educational context, as presumptively acceptable.[245]

Judges have paid special attention to two of the four fair use factors: (1) whether a use is transformative in purpose, and (2) how a use affects the market, or potential market, for an original work.[246] An analysis of

---

at a cost: people often forgo the original published articles for the summary excerpt and original link that Google and its competitors provide.").

239. Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 349 (1991).

240. *See generally* C.T. Drechsler, *Extent of Doctrine of "Fair Use" Under Federal Copyright Act*, 23 A.L.R.3d 139 (1969).

241. *Id.*

242. *Id.*

243. Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 608 (2d Cir. 2006).

244. Arista Records, LLC v. Doe 3, 604 F.3d 110, 124 (2d Cir. 2010) (internal quotations omitted).

245. *But cf.* HOWARD B. ABRAMS, THE LAW OF COPYRIGHT § 15:42 (2017).

246. *See* Jennifer Pitino, *Has the Transformative Use Test Swung the Pendulum Too Far in Favor of Secondary Users?*, 56 ADVOCATE 26, 28–29 (2013) ("According to the data, courts today are virtually ruling 100% of the time that a transformative use is a fair use. This has in essence created a per se rule that transformative use is fair use."); *see also* Robert W. Haskin, Jr. & Julie A. Pace, *Photocopying Is Cheap—Copyright Infringement Is Not*, ARIZ. ATT'Y, Nov. 1991, at 17, 18 ("[The]

major United States copyright lawsuits indicates that, unlike other recent technological advances such as home video recorders, searchable web archives, or digital snippets of books, online translation programs fall outside the scope of the fair use doctrine in its current form. We argue that legislators should amend copyright law to carve out fair use exceptions to certain uses of machine translation online.

C. Are AI Translation Services or Their Users Indirect Infringers?

Are internet-related companies, such as Google, accountable for copyright infringement when they offer machine translation services online? Or are the users of these services at fault? Before internet use was widespread, courts still had to grapple with the ways in which technological advances could facilitate copyright infringement.[247] In particular, courts have repeatedly addressed the issue of secondary liability, whether in the form of contributory infringement—meaning that a defendant did not directly engage in the infringing activity but somehow promoted or induced it—or vicarious liability—meaning that a defendant derived a direct financial benefit from infringing activity and had the ability to interrupt it, but did not.[248] Secondary liability for copyright infringement assigns responsibility to those "who do not directly infringe copyright rights, but, who, as a matter of fairness or to serve some other overarching policy concern, should nonetheless be liable for the copyright infringement of others taking place within their ambit."[249] One oft-cited advantage of secondary liability is cost-effectiveness, as pursuing a single case against a secondary infringer is less expensive than pursuing many cases against direct, individual infringers.[250]

In *Sony v. Universal City Studios*, for example, the United States Supreme Court addressed the issue of whether manufacturers of home videotape recorders were guilty of contributory infringement, since many

---

most important factor" of the fair use doctrine is that "which focuses on the effect of the use upon the potential market for, or value of, the copyrighted work.").

247. *See, e.g.*, Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 58 (1884) (holding that "[w]e entertain no doubt that the constitution is broad enough to cover an act authorizing copyright of photographs, so far as they are representatives of original intellectual conceptions of the author").

248. *See* Gershwin Publ'g Corp. v. Colum. Artists Mgmt., Inc., 443 F.2d 1159, 1162 (2d Cir. 1971) ("One who, with knowledge of the infringing activity, induces, causes or materially contributes to the infringing conduct of another, may be held liable as a 'contributory' infringer . . . [and] one may be vicariously liable if he has the right and ability to supervise the infringing activity and also has a direct financial interest in such activities.").

249. Craig A. Grossman, *From Sony to Grokster, The Failure of the Copyright Doctrines of Contributory Infringement and Vicarious Liability to Resolve the War Between Content and Destructive Technologies*, 53 BUFF. L. REV. 141, 147 (2005).

250. *See* Lital Helman, *Pull Too Hard and the Rope May Break: On the Secondary Liability of Technology Providers for Copyright Infringement*, 19 TEX. INTELL. PROP. L.J. 111, 155–56 (2010).

owners of the videotape recorders were using them to copy and save television broadcasts, which are covered by copyright law.[251] The Court held that since some users were infringing and others were not, the appropriate question to ask was whether the home videotape recorders were "capable of commercially significant non-infringing uses."[252] After determining that they were, the Court applied the fair use factors and noted that although commercial uses of copyrighted material are presumptively unfair, "[a] challenge to a noncommercial use of a copyrighted work requires proof either that the particular use is harmful, or that if it should become widespread, it would adversely affect the potential market for the copyrighted work."[253] In this instance, the Court was not persuaded that some consumers' use of the videotape recorders to copy television programs was harmful to the market, and since use of the technology for substantial non-infringing purposes was possible, the fair use doctrine should apply.[254]

Nevertheless, Justice Blackmun penned a vigorous dissent, objecting to the majority's conclusion on the grounds that Congress had rejected a "private use" exemption to copyright infringement, observing that fair use does not apply "when a user reproduces an entire protected copyrightable work and uses it for its original purpose, with no added benefit to the public . . . [t]here is then no need whatsoever to provide the ordinary user with a fair use subsidy at the author's expense."[255]

Changes in the way consumers could access music also raised new copyright problems.[256] In 2001, the Ninth Circuit heard the case of *A&M Records, Inc. v. Napster, Inc.*[257] A&M Records, a company that recorded, distributed, and sold music, had sued Napster for contributory and vicarious copyright infringement.[258] Napster was a start-up business that facilitated peer-to-peer file sharing, allowing its users to search for other users' files through a centralized index (a kind of electronic library card catalog) before copying and saving them on their own computers.[259] Since Napster itself was not doing the file copying, and was merely providing

---

251. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 422–23, 434–42 (1984).

252. *Id.* at 442.

253. *Id.* at 451.

254. *Id.* at 455.

255. *Id.* at 472, 480.

256. *See also* Joshua P. Binder, *The Future of Streaming Technology After Grokster*, L.A. L., Dec. 2005, at 13 (noting that "several technological changes have converged to cause a sea change in the way musical recordings can be delivered to consumers"). *See generally* Grossman, *supra* note 249 (discussing the challenges of copyright enforcement for online service providers).

257. A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004 (9th Cir. 2001).

258. *Id.* at 1010–11.

259. *Id.* at 1011.

users with the centralized index and allowing them to connect with one another, the court needed to address the issue of secondary liability.[260]

On the question of contributory infringement, the court examined whether Napster knew or had reason to know of acts of infringement by its users and whether it was materially contributing to such infringement; because Napster was aware of user downloads of copyrighted music files and because it facilitated such downloads through its centralized index, the court found Napster was a contributory infringer.[261] On the issue of vicarious liability for infringement, the court asked whether Napster had the ability to supervise its users and whether it benefited financially from their infringement.[262] Here, too, the court reasoned against Napster, finding that it was vicariously liable for infringement because it was materially benefiting from an infringement that it could have prevented.[263] As a result of this lawsuit, Napster was effectively out of business,[264] but its demise led to the rise of music industry-supported commercial services such as iTunes, which capitalize on user demand for single songs rather than entire albums without infringing on artists' copyrights.[265]

In 2005, the Supreme Court heard another music industry case: *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*[266] Grokster was a company that marketed itself as a successor to Napster, but it operated without a centralized server; individual users of the service contacted each other regarding what content they were willing to share using Grokster's Gnutella technology.[267] The Court was not impressed, holding that "one who distributes a device with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement, is liable for the resulting acts of infringement by third parties."[268] Whereas in the *Sony* case, the Court could not attribute specific

---

260. *Id.* at 1019, 1022.

261. *Id.* at 1022.

262. *Id.* at 1024.

263. *Id.*

264. *See Napster, The Long Goodbye*, ECONOMIST (May 16, 2002), http://www.economist.com /node/1134869 [https://perma.cc/P5RX-5GKS] (reflecting on the demise of Napster as both "the triumph of fat cats over humble music geeks" and a turning point in how consumers purchase music).

265. *But see* Sydney A. Beckman, *The Copyright Crusher: How Apple's iTunes Is a Vehicle Designed for Copyright Infringement and Apple's Legal Liability from Its Creation*, 1 TEX. A&M L. REV. 901, 909 (2014) (arguing that because iTunes allows the conversion of compact discs to a format that works with various music players, it violates copyright holders' exclusive rights).

266. Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd., 545 U.S. 913 (2005).

267. *Id.* at 913.

268. *Id.* at 919; *see also* Mark A. Lemley & R. Anthony Reese, *Reducing Digital Copyright Infringement Without Restricting Innovation*, 56 STAN. L. REV. 1345, 1346–47, 1362–65 (2004). The authors note that "[s]uing actual infringers is becoming passé in digital copyright law. In the digital environment, the real stakes so far have been in suing those who facilitate infringement by others." *Id.* at 1346.

knowledge of infringement to Sony and was unwilling to impute such knowledge to the company; in *Grokster*, the defendant company's own statements about its potential uses were evidence of vicarious infringement.[269]

Collectively, these cases illustrate the attention United States courts have paid not only to instances of direct copyright infringement, but also to more indirect circumventions of the spirit of the law, which are often made easier by emerging technologies. By implication, these cases raise the following questions: Are users who paste copyrighted material into machine translation programs and click the translate button infringers? Or is the provider of these services the infringer? In short, are online machine translation programs like Napster and Grokster? We think the parallels are strong, but as we discuss further in this Article, amendments to copyright law and the fair use doctrine could easily—and should—address the problems of copyright infringement in AI online translation to provide a more desirable outcome.

D. Purpose and Character: Transformative Use, Google Books and HathiTrust

The first fair use factor, which examines the "purpose and character" of a use, invites courts to consider whether a use "adds something new, with a further purpose or different character" and does "not substitute for the original use of the work."[270] An examination of United States case law indicates that online translation does not meet this definition of transformative.[271]

In 2004, Google started developing the Google Books Library Project.[272] Its goal was to scan the book collections of major libraries and make the digitized information available to both the libraries and to the general public, which would have access to randomized portions of the books via Google's search engine.[273] Under United States copyright law, copyright holders—such as the authors or publishers of books—have the exclusive right (for a limited time) to make and sell copies of their works, to create derivative works, and to perform or display their works publicly.[274] Although many of the works it was scanning were still in

---

269. Sony v. Universal City Studios, 464 U.S. 417, 439 (1984); *see also Grokster*, 545 U.S. at 939.

270. *More Information on Fair Use*, U.S COPYRIGHT OFF., https://www.copyright.gov/fair-use/more-info.html [https://perma.cc/DXK6-U3TQ].

271. *See, e.g.*, Authors Guild v. Google, Inc., 804 F.3d 202, 215 (2d Cir. 2015), *cert. denied sub nom.* 136 S. Ct. 1658 (2016).

272. *See* Genan Zilkha, *Fair Use: An Overview*, 86 N.Y. ST. B.J. 40, 43 (2014).

273. *Id.*

274. 17 U.S.C. § 102 (2018).

copyright, Google did not contact any authors or publishers as part of the Google Books Library Project. In 2005, the writers' advocacy group The Authors Guild accused Google of committing large-scale copyright infringement.[275]

Following two Second Circuit court decisions in favor of Google and the United States Supreme Court's denial of the Authors Guild's petition for certiorari, Google was free to pursue its Library Project under the fair use doctrine's four factors.[276] Ultimately, the courts held that the Google Books Library Project offered a transformative use of copyrighted works that did not usurp authors' rights to produce derivative works and did not significantly affect the potential market for the authors' works.[277] The Second Circuit held that use of digital copies to provide a search function "augments public knowledge by making available information *about* Plaintiffs' books without providing the public with a substantial substitute for matter protected by the Plaintiffs' copyright interests in the original works or derivatives of them."[278] The same was found true of Google's "snippet" function, which made small sections of original text visible.[279] Thus, the use of copyrighted material was transformative in that its purpose did not directly mirror that of the original authors. This analysis would appear to not apply to online translation services, which provide users with derivative works—translations—that are a substantial substitute for authors' copyright interests in their original works.

In a related case, *Authors Guild, Inc. v. HathiTrust*, authors and various authors' associations brought suit against an organization called HathiTrust.[280] Founded by a group of research universities to house digital copies of their libraries, which they wanted to share with one another, HathiTrust's Digital Library (HDL) was a repository of "digital copies of more than ten million works, published over many centuries, written in a multitude of languages, covering almost every subject imaginable."[281] The plaintiffs alleged that HathiTrust's digitalization of copyrighted works without author permission, and their conversion of works into formats for print-disabled individuals, were violations of the Copyright Act.[282] The court came down on the side of fair use, guided principally by the fact that

---

275. *See* Zilkha, *supra* note 272.

276. *See generally Fair Use: Mass Digitization of Books*, *in* LINDEY ON ENTERTAINMENT, PUBLISHING AND THE ARTS § 5:10.30 (3d ed. 2017) [hereinafter LINDEY ON ENTERTAINMENT, § 5:10.30] (discussing case law).

277. *Id.*; *see also* Authors Guild v. Google, Inc., 804 F.3d 202, 207–08 (2d Cir. 2015).

278. Authors Guild v. Google Inc., 804 F.3d at 207–08.

279. *Id.*

280. Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 92 (2d Cir. 2014).

281. *Id.* at 92.

282. *See id.*

the service did not add into circulation entire works; it merely allowed users to conduct "word searches" to locate specific phrases and page numbers in digitized books.[283] Thus, scanning the full works did not amount to taking an excessive portion of the copyrighted works. Further, the authors would not suffer economic injury because the service did not serve as a substitute for their original works.

In its analysis, however, the court made several relevant comments that suggest translations should be treated differently. To start, in its definition of what constitutes transformative use of a copyrighted work, the court noted that "[a] use is transformative if it does something more than repackage or republish the original copyrighted work."[284] The court disputed the notion that the use of digital copies to facilitate access for print-disabled persons was transformative, noting that:

> [A] transformative use adds something new to the copyrighted work and does not merely supersede the purposes of the original creation . . . . By making copyrighted works available in formats accessible to the disabled, the HDL enables a larger audience to read those works, but the underlying purpose of the HDL's use is the same as the author's original purpose. Indeed, when the HDL recasts copyrighted works into new formats to be read by the disabled, it appears, at first glance, to be creating derivative works over which the author ordinarily maintains control. As previously noted, *paradigmatic examples of derivative works include translations of the original into a different language, or adaptations of the original into different forms or media . . . .* It is true that, oftentimes, the print-disabled audience has no means of obtaining access to the copyrighted works included in the HDL. But, *similarly, the non-English speaking audience cannot gain access to untranslated books written in English and an unauthorized translation is not transformative simply because it enables a new audience to read a work.*[285]

While the court held that providing access to the print-disabled was not a transformative fair use of copyrighted material, it found that such a use was nevertheless a valid purpose supported by Supreme Court decisions and statements from Congress, which carved out special exceptions to copyright law to benefit the disabled. This suggests that, absent congressional amendments to the Copyright Act, an automated online translation of a copyrighted work would not be considered transformative under the first fair use factor (purpose). The fact that it

---

283. *Id.* at 98 ("Because it was reasonably necessary for the HDL to make use of the entirety of the works in order to enable the full-text search function, we do not believe the copying was excessive.").

284. *Id.* at 96.

285. *Id.* at 101 (citations omitted and emphasis added).

enables a larger audience to read the work is insufficient to satisfy this branch of the fair use doctrine.

E. Commercial Impact of Use

Under the fourth fair use factor, United States courts have also addressed the commercial impact of unauthorized translations by examining "whether, and to what extent, the unlicensed use harms the existing or future market for the copyright owner's original work."[286] In evaluating this factor, courts "consider whether the use is hurting the current market for the original work (for example, by displacing sales of the original) and/or whether the use could cause substantial harm if it were to become widespread."[287] Because unauthorized translations compete with authorized translations and because poor quality translations may harm an author's moral rights, online translation programs are unlikely to meet the test for fair use under the fourth factor.

United States courts have had occasion to examine fair use defenses to unlicensed translations, looking at the impact they have on the market for authors' works. For example, in 1987, in *Radji v. Khakbaz*, a dispute concerning the unauthorized translation of excerpts of an author's work published in an Iranian newspaper—after the same excerpts had appeared with the author's permission in a British newspaper—a district court judge in the District of Columbia held that the fair use defense did not apply, largely because "[d]efendants' reprinting of the 80 days of diary entries in a commercial publication was presumably intended to boost sales and thus was for commercial rather than non-profit educational purposes."[288]

In 1999, in *Nihon Keizai Shimbun, Inc. v. Comline Business Data, Inc.*, a Japanese publisher of financial, business, and industry news brought a copyright and trademark infringement action against a United States company and its officers based on the company's sale of roughly translated "abstracts" of the publisher's news articles.[289] The court held that the defendant was guilty of infringement because the company's news abstracts precisely replicated the original publisher's structure and organization of facts.[290]

> The Comline abstracts appear to be direct, if not word-for-word, translations of the Nikkei articles, edited only for clarity. The average Comline abstract uses about two-thirds of the protectible material in

---

286. U.S. COPYRIGHT OFF., *supra* note 270.

287. *Id.*

288. Radji v. Khakbaz, 607 F. Supp. 1296, 1300 (D.D.C. 1985), *amended*, No. 84-0641, 1987 WL 11415 (D.D.C. May 15, 1987).

289. Nihon Keizai Shimbun, Inc. v. Comline Bus. Data, Inc., 166 F.3d 65, 69 (2d Cir. 1999).

290. *Id.* at 73.

the corresponding Nikkei article. The abstracts track the information
in the articles sentence by sentence, in sequence; only occasionally
do the abstracts combine two Nikkei sentences, divide a sentence, or
rearrange the facts among different sentences. Comline adopts, by
and large, the exact same structure and organization of the facts
reported by Nikkei.[291]

Though careful to highlight that facts are not copyrightable, the court
here focused on the sale of abstracts that were essentially replicas of
copyrighted news articles, translated quickly by human employees.[292] The
actions taken by the defendants in both of these cases, except for paying
at least some attention to quality and abstracting the content, are highly
similar in their commercial nature to those of automated translating
systems.[293] Though automated translation providers may not directly
charge users for translations, they are for-profit companies that are able to
sell advertising and user data, making their use of authors' works
commercial.

F. The Fair Use Doctrine Does Not Cover AI Translation—But Maybe It Should, Sometimes

AI translation online presents an unusual set of copyright issues that
United States courts have not yet directly addressed. However, based on
case law focused on other technological innovations, it appears that the
current fair use doctrine would not apply to the machine translation of
copyrighted material online. These examples illustrate tensions similar to
those that arise with the proliferation of translation tools online. Authors
have the right to make copies and create derivative versions of their works,
and according to the Berne Convention, this means they have the "right to
translate" their works; yet automated online translation systems do not
seek permission from authors to translate their works.[294] Further, while
Sony home video recorders faithfully reproduced original television

291. *Id.* at 71.
292. *Id.*
293. *See* Douglas L. Rogers, *Increasing Access to Knowledge Through Fair Use—Analyzing the Google Litigation to Unleash Developing Countries*, 10 TUL. J. TECH. & INTELL. PROP. 1, 63 (2007) ("[In] both Radji and Nihon . . . the translations of defendants were sold commercially and were substitutes for the translations of the plaintiffs.").
294. *See, e.g.*, *Google Terms of Service*, GOOGLE, https://www.google.com/intl/en/policies/terms [https://perma.cc/MK2A-Z22N]. The company states that it provides its services "as is," and makes no warranties about "the content within the services, the specific functions of the services, or their reliability, availability, or ability to meet your needs." *Id.* The terms also instruct users not to misuse Google services, and note that use of the services does not give users ownership of any intellectual property rights over content they may access. *Id.* However, Google *itself* is using copyrighted material for profit. *See id.*

content and Napster facilitated the sharing of existing songs, translation services often produce distorted, poor quality translations of original works which may harm authors' reputations.[295] Like Sony, translation services could argue that they merely provide users with a tool, similar to a video recorder or a camera, and that they cannot be held responsible for any infringing uses of the tool. However, end-users of these services are accessing complete original works in unauthorized derivative form, and such use is inherently infringing, rather than substantially non-infringing, and services such as Google Translate are fully aware of the fact because they explicitly offer to translate entire pages of web content.[296] In this sense, translation services closely resemble Napster or Grokster, as they know of infringing uses of their platforms, encourage rather than limit them and also benefit financially from them.

The way the *HathiTrust* court applied the fourth fair use factor, which looks at the market impact of a contested use of a copyrighted work, also indicates that translations could be protected in a way that the page numbers and keyword searches of Google Books are not because translation services may undermine authors' reliance on quality translations for their works. The *HathiTrust* court held that the fair use "doctrine is generally subject to an important proviso: A fair use must not excessively damage the market for the original by providing the public with a substitute for that original work."[297] For example, a book review that quotes from a book is acceptable as fair use both because it is not a substantial portion of the work in question and because—even if the review is negative and discourages readers from buying the book, causing economic harm to the author—the review is not a substitute for the book.[298] Fair use also allows an internet search engine to "display low-resolution versions of copyrighted images in order to direct the user to the website where the original could be found."[299] Translations of web pages

---

295. *See* STEINER, *supra* note 26; *see also* FREEDMAN, *supra* note 25.

296. *See, e.g.*, GOOGLE TRANSLATE, https://translate.google.com [https://perma.cc/HDA3-V4XP]. The company invites users to automatically view any web page in their preferred language. *Id.*; *see also* BING TRANSLATE, https://www.bing.com/translator [https://perma.cc/DB8V-S9UC]. Microsoft, the progenitor of Bing Translate, also invites users to "[i]nstantly translate [their] text from one language to another with Translator for Bing." *Try Microsoft Translator for Free*, MICROSOFT TRANSLATOR, https://www.microsoft.com/en-us/translator/business/trial/ [https://perma.cc/KP5W-GV2U]; *see also Machine Translations for Personal, Business and Enterprise Usage*, SYSTRAN, http://www.systransoft.com/lp/machine-translation [https://perma.cc/FS8V-7C3J]. The company explains that "[the software] lets you understand foreign language information—like any Web page, document or email—in the language of your choice in real-time." *Machine Translations for Personal, Business and Enterprise Usage*, SYSTRAN, http://www.systransoft.com/lp/machine-translation [https://perma.cc/FS8V-7C3J].

297. Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 95 (2d Cir. 2014).

298. *Id.*

299. *Id.*

or articles are often not mere quotes or the equivalent of low-resolution thumbnails, but they are the entirety of a copyrighted work,[300] and reading such a translation may substitute a search for the original work or an authorized translation of it—especially as the translation services do not direct users to alternate sources with better-quality translations.

Finally, the *HathiTrust* court cited the importance of mass download prevention systems when it noted that even where the libraries involved in the case permitted works to be read online, they made efforts to ensure that "inappropriate levels of access" did not take place.[301] Currently, no such barriers prevent automated services from translating copyrighted material online.

However, legislators could easily change this by carving out a special set of fair use factors that apply to AI translation online. These could specify that automated translation of certain material on the web is non-infringing even when it covers the entire protected copyrightable work. For instance, fair use could be extended to the automated translation of "snippets" of text on the web, sufficient to get the gist of content without reproducing it entirely. This approach would mirror that of the Second Circuit in the Google Books Library Project, which focused on the amount and substantiality of copyrighted works being made available to internet users.[302] It would allow the *de minimis* translation of copyrighted material useful to people conducting research online while prohibiting the wholesale translation of lengthy news articles without the publishers' consent. Fair use could also consider damage to the markets for an original work and for a translation of it, looking at the impact of an automated translation on an author's ability to profit from his work as well as any reputational damage arising from a poor translation, echoing the concerns of *Radji* and *Nihon*.[303] In many instances, where the amount of text translated is minimal and for private use, AI web translations would easily clear this hurdle.

Challenges to these explicit limitations are likely to arise from those who argue that copyright law in general has grown too expansive, at the expense of broad public access to knowledge.[304] Yet failure to protect

---

300. *See* Ketzan, *supra* note 47, at 233. The author notes that the fair use doctrine could cover machine translation "were it not for the fact that it translates entire Web pages." *Id.*

301. *Authors Guild, Inc.*, 755 F.3d at 100 (citing the HathiTrust Digital Library Executive Director).

302. *See* LINDEY ON ENTERTAINMENT, *supra* note 276, at § 5:10.30.

303. *See* Nihon Keizai Shimbun, Inc. v. Comline Bus. Data, Inc., 166 F.3d 65 (2d Cir. 1999); Radji v. Khakbaz, 607 F. Supp. 1296, 1300 (D.D.C. 1985), *amended* No. 84-0641, 1987 WL 11415 (D.D.C. May 15, 1987).

304. *See, e.g.*, Jenny Lynn Sheridan, *Copyright's Knowledge Principle*, 17 VAND. J. ENT. & TECH. L. 39 (2014).

authors' rights can also diminish public access by discouraging creators from sharing their works in the first place.[305] An analysis of the copyright issues of translation cannot ignore that the companies providing translation on the web are for-profit multinationals. While international copyright law creates exceptions for educational or research purposes, these special cases do not extend to the unauthorized use of copyrighted material for financial gain.[306]

Lawmakers should act to clarify which uses of AI online translation qualify as fair use to preserve copyright law while also facilitating the public's access to culture and knowledge.

G. Comparative Perspective: Ideas from Abroad

While the United States, the United Kingdom, Canada, and France all recognize copyright law and grant authors limited monopolies over their works, each of these jurisdictions also carves out exceptions, creating limited circumstances under which use of an author's material without permission does not constitute copyright infringement.[307] "From its beginning, the law of copyright has developed in response to significant changes in technology. Indeed, it was the invention of a new form of copying equipment—the printing press—that gave rise to the original need for copyright protection."[308] Because AI translation platforms are global, the theoretical justifications for copyright law in different countries provide an important means for analyzing this new technological development and its creation of copyright disputes.

1. Copyright in the United Kingdom, Canada, and France

In its emphasis on protecting the property rights of original works produced by creative labor, the United Kingdom "approach fits fairly neatly with John Locke's labour theory as a justification of property rights."[309] The United Kingdom's Copyright, Designs and Patents Act 1988 protects "(a) original literary, dramatic, musical or artistic works, (b) sound recordings, films [or broadcasts], and (c) the typographical arrangement of published editions" and grants all authors the right of

---

305. *See, e.g.*, Thomas M. Gage, *Whelan Associates v. Jaslow Dental Laboratories: Copyright Protection for Computer Software Structure—What's the Purpose?*, 1987 WIS. L. REV. 859, 887 (1987) (arguing that copyright for computer programs advances the public welfare and "copyrights encourage the development of new programs by allowing owners to exclude free riders and recover developmental costs").

306. *See* Berne Convention, *supra* note 82, at arts. 10 and 10*bis*.

307. *See generally* 17 U.S.C. § 107 (2018).

308. Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417, 430 (1984).

309. ESTELLE DERCLAYE, RESEARCH HANDBOOK ON THE FUTURE OF EU COPYRIGHT 288 (2009).

reproduction or the right to control how and when copies of their works are made.[310] Depending on the type of work at issue, the UK also grants authors the distribution right, the rental right, the public performance right, the communication right, and the adaptation right.[311]

Like the United States, the United Kingdom allows some exceptions to copyright, known as "fair dealing" exceptions. These, however, are more limited than their United States counterparts.[312] Fair dealing allows some use of copyrighted material for purposes of non-commercial research or study, criticism and review, or news reporting.[313] To use this defense, the claimed infringer must show not only that the copying falls into one of these three categories, but also that it is "fair" and often, that it contains sufficient acknowledgement for the original author.[314] Factors when considering the "fairness" of the copying can include the quantity of the work taken, the infringer's motives, and how the infringement will affect the author's returns for the copyrighted work.[315] In 2014, the United Kingdom introduced a new fair dealing exception, allowing some text and data mining as long as these were done for non-commercial research purposes.[316] Copyright law in the United Kingdom typically lasts for "70 years from the end of the calendar year in which the last remaining author of the work dies."[317]

Canadian copyright law was passed with the Copyright Act of Canada in 1921 and amended in 1988, 1997, and 2012.[318] The Act covers "every original literary, dramatic, musical and artistic work," but also performances, sound recordings and communication signals,[319] and gives authors copyright —which covers the right to "produce, reproduce, perform or publish any translation of the work"[320]—moral rights and "neighbouring" rights. Moral rights include attribution, integrity, and association, allowing authors to have some control over what their work is associated with, while neighbouring rights pertain to the performance, transmission, and reproduction of creative works and are especially relevant in the music industry.[321] Fair dealing exceptions cover "research,

---

310. The Copyright, Designs, and Patents Act 1988, c. 1 (UK).

311. *Id.* at c. 2.

312. *Id.* at c. 3.

313. *Id.*

314. *Id.*

315. *Id.*

316. *Id.*

317. *Id.*

318. *See* Copyright Act, R.S.C. 1985, c. C-42 (Can.).

319. *Id.*

320. *Id.*

321. *See Royalties, Neighbouring Rights, Private Copying Levy*, MUSICIANS' RTS ORG. CAN., https://musiciansrights.ca/en/royalties/ [https://perma.cc/9RQ6-DDKM].

private study, education, parody or satire," "criticism and review," and "news reporting."[322] There is also an exception for "non-commercial user-generated content," which allows some use of materials available to the public subject to the following exceptions:

> (a) the use of, or the authorization to disseminate, the new work or other subject-matter is done solely for non-commercial purposes;

> (b) the source — and, if given in the source, the name of the author, performer, maker or broadcaster — of the existing work or other subject-matter or copy of it are mentioned, if it is reasonable in the circumstances to do so;

> (c) the individual had reasonable grounds to believe that the existing work or other subject-matter or copy of it, as the case may be, was not infringing copyright; and

> (d) the use of, or the authorization to disseminate, the new work or other subject-matter does not have a substantial adverse effect, financial or otherwise, on the exploitation or potential exploitation of the existing work or other subject-matter — or copy of it — or on an existing or potential market for it, including that the new work or other subject-matter is not a substitute for the existing one.[323]

Canada has two official languages, English and French, and the Official Languages Act requires the federal government to provide services to citizens in both languages.[324] Canada is also a Crown copyright country, meaning special copyright rules cover the distribution of state-produced materials.[325] Copyright in Canada lasts for the lifetime of an author plus fifty years.[326]

The first article of the French Intellectual Property Code is unambiguous in emphasis: it states that "*[l]'auteur d'une oeuvre de l'esprit*" (the author of a *work of the spirit*) by virtue of its creation alone, enjoys an exclusive property right in the work.[327] French *droit d'auteur*

---

322. Copyright Act, R.S.C. 1985, c. C-42 § 29 (Can.).

323. *Id*.

324. *Frequently Asked Questions*, CAN. OFF. OF THE COMM'R OF LANGUAGES (last modified NOV. 2, 2011), http://www.ocol-clo.gc.ca/en/resources/frequently-asked-questions [https://perma.cc/3ARA-29BS].

325. *Crown Copyright Request, When to Apply for Crown Copyright Clearance*, GOV. OF CAN. PUBLICATIONS (last modified Oct. 24, 2018), https://www.canada.ca/en/canadian-heritage/services/crown-copyright-request.html [https://perma.cc/V3LH-FRBP]. In the United States, by contrast, section 105 of the Copyright Act states that federal government works are in the public domain. *Policies: Public Domain & Copyright Notice*, U.S. GOV'T PUB. OFF., https://www.gpo.gov/help/public_domain_copyright_notice.htm [https://perma.cc/5JAR-WB3P].

326. Canada Copyright Act, R.S.C. 1985, c. C-42, § 12 (Can.).

327. CODE DE LA PROPRIETE INTELLECTUELLE [C. PROPR. INTELL.] [INTELLECTUAL PROPERTY CODE] art. L111-1 (Fr.) (emphasis added).

(author's rights) protects any original work, expressed in any form; thus, unlike the United States and UK, France does not require the work to be recorded in a tangible medium of expression, and covers, for example, original choreography.[328] The Code defines two distinct sets of rights for authors: *droits patrimoniaux* (proprietary rights) and *droits moraux* (moral rights).[329] Once an author's work has been published, French law carves out limited exceptions for non-infringing uses of the work, ranging from private family performances to research, press critiques, diffusion of current news, and parody, among others.[330] Copyright in France lasts for an author's lifetime plus seventy years,[331] with some variation for posthumous works or those by authors who died in battle defending France.[332]

In each of these jurisdictions, copyright law strives to protect authors' rights while recognizing that there are limited circumstances in which the public's use of original material should not be considered infringing. Technological advances often pose legal challenges as different jurisdictions seek to maintain this balance.[333]

2. Web Innovation and Copyright Law Outside the United States

An examination of internet copyright issues from other jurisdictions suggests that there, too, online translation would likely fall outside the parameters of fair dealing or other exceptions to infringement. In the United Kingdom and Canada, problems arise under a fair dealing analysis because the non-commercial purpose of the translation would not be satisfied: the companies providing the translations are commercial enterprises.[334] In addition, the amount of material such programs copy is currently unlimited, and while it is unclear how translation could affect the market for individual authors' works, it could conceivably harm the market for good-quality translations, which would violate fair dealing

---

328. *Id.* at art. L112-2.

329. *Id.* at arts. L111-1, L121-1, L122-1.

330. Société des Gens de Lettres, *Les exceptions au droit d'auteur*, https://www.sgdl.org/juridique/la-minute-juridique-de-la-sgdl-sur-webtv/758-les-exceptions-au-droit-dauteur (last visited Nov. 27, 2017).

331. CODE DE LA PROPRIETE INTELLECTUELLE [INTELLECTUAL PROPERTY CODE] art. L123-1 (Fr.).

332. *Id.* at arts. L123-4, L-123-10.

333. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 433 (1984).

334. *See* Michael Geist, *Why Copyright Law Poses a Barrier to Canadian AI Ambitions*, GLOBE & MAIL (May 17, 2017), https://beta.theglobeandmail.com/report-on-business/rob-commentary/why-copyright-law-poses-a-barrier-to-canadian-ai-ambitions/article35019241/?ref=http://www.theglobeandmail.com [https://perma.cc/RTG7-CETG]. Geist, a Canadian professor specialized in internet and e-commerce law, argued that the country's copyright laws "may hamper the ability of companies and researchers to test and ultimately bring new AI services to market." *Id.*

principles.[335] Further, unlike internet service providers, which the Supreme Court of Canada held are not users under the country's Copyright Act[336] (and therefore do not have to pay royalties for illegally downloaded music), translation programs would likely fall under sections 2.4(1) and 2.4(1.1) of the Act, which cover "communication to the public by telecommunication" and provide that "communication of a work or other subject-matter to the public by telecommunication includes making it available to the public by telecommunication in a way that allows a member of the public to have access to it from a place and at a time individually chosen by that member of the public."[337] In France, which places significant emphasis on moral rights, translations online may face barriers not only for their infringement on the rights of reproduction and translation but also because they deprive authors of the right to preserve the integrity of their works.[338]

Many European Union publishers have voiced dissatisfaction with the relative lack of regulation of major search engines, and "[s]ome E.U. member states have responded . . . by implementing legislation that requires search engines to pay a tax to publishers for displaying these excerpts on their search results page."[339] Spain, for instance, passed legislation in 2015 requiring Google to pay publishers a tax if it wanted to provide links to the publishers' content.[340] Writing in the *Temple International and Comparative Law Journal*, Christopher Gagne argued that such a move would help counteract the exploitation by search engine powerhouses of publishers' works for commercial gain, which tends to "[eliminate] fair competition within the industry, and [remove] the incentive to innovate among publishers."[341] By creating a tax on the publication of article content, Gagne suggested that legislation, such as that introduced in Spain, would "eliminate search engines' ability to leverage their dominant position within the digital market against publishers' rights to compensation for reproduction of their original works."[342]

---

335. *See* The Copyright, Designs, and Patents Act 1988, c. 1 (UK).

336. Soc'y of Composers, Authors, & Music Publishers of Can. v. Can. Ass'n of Internet Providers, [2004] 2 S.C.R. 427 (Can. S.C.C.); *see also* Julie DeWeese, *Legislating Against Copyright Infringement Across Borders: An Examination of Canada's Efforts to Keep Pace with Technology*, 17 L. & Bus. Rev. Am. 347 (2011).

337. Canada Copyright Act, R.S.C. 1985, c. C.-42.

338. *See* William F. Patry, *Rationale for Droit Moral*, *in* 5 Patry on Copyright § 16:2 (2017).

339. Gagne, *supra* note 238, at 204.

340. *Id.*

341. *Id.*

342. *Id.*

The Spanish legislation is also indicative, however, of a generalized European distrust of the monopoly Google, in particular, has established online, and a resistance to the notion that because web aggregators may provide additional access to original content, they should be immune from claims of copyright infringement.[343] We argue that copyright protection should not "prevent the normal functioning and the development of new technologies, but . . . it is necessary to ensure an adequate protection of copyright in the information society. Technological development allows, in fact, faster and easier reproduction of works, for which reason protection of copyright must adapt to this technological development."[344] Further, as the internet assumes an ever-more-dominant role in the way people communicate with one another, European countries have questioned the power of companies facilitating such communication, expressing concerns that these large firms have "so far slipped through the regulatory net, operating without burdens and without regulation."[345]

Numerous European countries reacted unfavorably to the 2005 Google Books ruling in the United States,[346] and it is unlikely that a similar result would have emerged in the European Union.[347] In 2009, French publishing group La Martinière and others similarly situated won a lawsuit against Google for copyright infringement, after Google included books from the publishing group in its Google Books Library Project without permission.[348]

Specifically, La Martinière alleged a violation of its right of reproduction, copyright infringement, trademark infringement, and unfair competition.[349] The French publishers' union (*Syndicat national de*

---

343. *See* U.S. CONST. art. I, § 8, cl. 8.

344. *See* Case C-5/08, Infopaq Int'l v. Danske Dagblades Forening, 2009 E.R.C. I-06569 ("The storing and subsequent printing of an extract from a newspaper article which contains the search word and the five words which precede and follow it must be considered as a reproduction within the meaning of Article 2 of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.").

345. *See* Emily B. Laidlaw, *Private Power, Public Interest: An Examination of Search Engine Accountability*, 17 INT'L J.L. & INFO. TECH. 113, 114 (2009).

346. Kevin J. O'Brien & Eric Pfanner, *Europe Divided in Google Book Deal*, N.Y. TIMES (Aug. 23, 2009), http://www.nytimes.com/2009/08/24/technology/internet/24iht-books.html [https://perma.cc/W7T9-XMGY].

347. Raquel Xalabarder, *Google Books and Fair Use: A Tale of Two Copyrights?*, 5 J. INTELL. PROP., INFO. TECH. & ELECTRONIC COMM. L. 53, 56 (2014) ("It is very unlikely that a similar result could be achieved under any EU law. Fair use does not exist in Europe on a general basis.").

348. Tribunal de grande instance Paris, Dec. 18, 2009, SAS Éditions du Seuil et a. c/ Sté Google Inc., No. 09/00540.

349. Olivier Chicheportiche, *Droit d'auteur: Google Contre-attaque, La Martinière Réclame 15 Millions d'euros*, ZDNET (Sept. 25, 2009), https://www.zdnet.fr/actualites/droit-d-auteur-google-contre-attaque-la-martiniere-reclame-15-millions-d-euros-39707865.htm [https://perma.cc/E7ZN-72V8].

l'édition) and the *Société des gens de lettres* (SGDLF), a writers'
association founded in 1838 by Honoré de Balzac, Victor Hugo,
Alexandre Dumas and George Sand, had joined the lawsuit in support of
La Martinière.[350] The French court's finding of copyright infringement
prevented Google from including French works that were still in copyright
in its Books project, although it was able to draw on works for which the
copyright had expired.[351] Though Google initially appealed the ruling, in
2011 it signed an agreement with La Martinière covering out-of-print
editions, and this deal ended the legal dispute between the companies.[352]
Such agreements appear to be a common compromise: in 2013, Google
and French government officials struck a deal that would quash proposed
legislation similar to the Spanish tax, on the condition that Google support
the digitization of French publications through major financial
investments.[353]

In Europe, as in the United States, courts continue to grapple with
the best way to balance copyright protection with technological and—in
particular—internet innovation. In 2016, as part of its proposed update of
European Union copyright rules, the European Commission suggested a
new copyright exception that would allow research organizations to use
text and data mining tools to analyze digital content that has been lawfully
acquired or accessed.[354] This suggests that any fair use-type exceptions to
translation online may be palatable to European courts only in a non-
commercial, research-driven setting. Because AI online translation
platforms reach a global audience, United States lawmakers must work
with their counterparts in Canada, Europe, and around the world in closing
the gaps of today's copyright law, which must be updated to reflect new
technological advances.

VI. Crafting a More Balanced Solution

Can copyright law address the unusual problems posed by AI
translation online? Online translation services continue to suffer from
inherent issues as to quality and bias—issues which can ultimately deform

---

350. SOCIETE DES GENS DE LETTRES, https://www.sgdl.org [https://perma.cc/87GB-76FW].

351. *See* Tribunal de grande instance Paris, Dec. 18 2009, SAS Éditions du Seuil et a. c/ Sté
Google Inc., No. 09/00540.

352. *Google et La Martinière Trouvent un Accord sur la Numérisation des Livres*, LE MONDE,
(Aug. 25, 2011), https://www.lemonde.fr/technologies/article/2011/08/25/google-et-la-martiniere-
trouvent-un-accord-sur-la-numerisation-des-livres_1563287_651865.html [https://perma.cc/J87T-
SMJA].

353. Gagne, *supra* note 238, at 215.

354. *Commission Proposes Copyright Exception for Researchers*, EUR. COMMISSION (Sept. 14,
2016), http://ec.europa.eu/research/index.cfm?pg=newsalert&year=2016&na=na-140916 [https://
perma.cc/F72J-S54D].

an author's original work. In addition, the proliferation of cheap, poor-quality translations can effectively harm or destroy the market for the underlying work itself, as well as for good translations in general.[355]

On the other hand, online translations—especially into languages that are not widely spoken—can bring news and ideas to places that would never otherwise have access to them. In addition to broadening people's linguistic horizons, these advanced technological tools have a place in the public sphere where precision and style are of secondary importance and what really matters is getting a taste of another culture.[356] There is a market for AI translation, and it would be short-sighted to ignore and create meaningless legal hurdles for users.

This Article advocates greater balance to minimize the harms of AI translation services while maximizing their benefits. In its current form, the copyright regime cannot withstand the rapid developments of AI and online machine translation tools. Given the wide-scale use of Google Translate and its kin, enforcing existing copyright restrictions without making any adjustments would be a futile and inefficient means of balancing the public and private interests at the heart of intellectual property law justifications, whichever justification one adopts. We cannot stop technological progress; nor should we try.

Instead, it is time to take practical measures by combining legal and digital tools to address the copyright problems of AI translation online. The ultimate goal should be to continue promoting authors' rights while enabling access to knowledge and to cultural assets. New solutions must function across jurisdictions and stand the test of time and evolving technology.[357]

Currently, in the absence of rules specifically addressing online translation, the law ignores the copyright violations it causes. Instead, the law favors the interests of big companies such as Google and Microsoft, which have almost unlimited access to copyrighted materials to mine for profit by engaging with users and selling related advertising.[358] Although consumers can use online translation services without pulling out their

---

355. *See* Ketzan*, supra* note 47.

356. *See id.* at 233 (noting that online machine translation of a work may expand the market for it if the translation "piques the curiosity" of speakers of other languages).

357. *See, e.g.*, Shira Perlmutter, *Participation in the International Copyright System as a Means to Promote the Progress of Science and Useful Arts*, 36 LOY. L.A. L. REV. 323, 325 (2002) ("It is increasingly clear that the copyright balance, and its effect on incentives and the public interest, cannot be confined to a purely domestic sphere.").

358. *See* GOOGLE SAFETY CTR., https://safety.google/privacy/data [https://perma.cc/5DRK-K7JX]; *see also* Mark Hachman, *The Price of Free: How Apple, Facebook, Microsoft and Google Sell You to Advertisers*, PCWORLD (Oct. 1, 2015), https://www.pcworld.com/article/2986988/privacy/the-price-of-free-how-apple-facebook-microsoft-and-google-sell-you-to-advertisers.html [https://perma.cc/42FV-4QSX].

wallets, they are not receiving a free product: they (we) are paying with their (our) data. In 2017, for example, news reports indicated that Translate.com, which provides a free machine translation service powered by Microsoft Translator, had indexed a large volume of highly sensitive material its users had been translating, from termination letters to a global bank's staff performance report.[359] The company's Terms & Conditions stated that Translate.com could not and did not "guarantee that any information provided to [them] by you will not become public under any circumstances."[360] It further provided that "[y]ou should appreciate that all information submitted on the website might potentially be publicly accessible."[361]

In the next sub-section, we will describe our new proposed framework. We argue for balance between the conflicting goals at issue and suggest that balance is best achieved with a combination of digital and legal tools. We need to make more room for authors' rights without shutting down AI translation programs, which is possible by setting limits on the amount and type of content that may be translated automatically online and by requiring software designers to adopt a "fair use by design" approach to their AI translation systems that implements legal guidelines.[362]

Our framework is based on several conclusions drawn from the above discussions.

First, the law requires greater consideration for authors' rights in translations of their works—and in quality translations in general—but not at the expense of web-based cross-cultural exchange in the process. Second, legislators should generally allow the free use of AI translation programs, but clearly delineate the amount and type of content that may be translated automatically online and allow authors to opt out of having their webpages available for automatic machine translation. Third, AI translation platforms should adhere to fair use guidelines by design. Finally, policymakers should set guidelines establishing the reach of the fair use doctrine in AI translation, stating explicitly what constitutes an infringement, and WIPO should join this effort, devising its own norms in collaboration with policy makers internationally to ensure the rules' global

---

359. Florian Faes, *Translate.com Exposes Highly Sensitive Information in Massive Privacy Breach*, SLATOR (Sept. 7, 2017), https://slator.com/technology/translate-com-exposes-highly-sensitive-information-massive-privacy-breach [https://perma.cc/4J6Q-72CP].

360. *Id.*

361. *Id.*

362. *See* Niva Elkin-Koren, *Fair Use by Design*, 64 UCLA L. REV. 1082, 1085 (2017) (arguing that "for fair use to serve its role in the twenty-first century, the checks that it intends to create on the rights of authors must also be embedded in the design of online systems").

applicability. By reducing uncertainty, this framework will protect all of the interests at the core of intellectual property law.

A. First Conclusion: It's Time for New Fair Use Guidelines and Digital Traffic Lights

The law should respect the rights of authors to control use of their works—and not give large multinational corporations a free pass to use any work which is accessible to the public—but it should accomplish this aim without impeding technological developments. As a first step, lawmakers should delineate what amount and, under what circumstances, AI translation constitutes fair use.

In his 2007 article examining the copyright issues of translation, Eric Ketzan discussed creating an affirmative "you may translate" metatag, which translation providers could offer authors as an explicit license.[363] Metatags are behind-the-scenes HTML code on webpages. Although they would not entirely prevent unauthorized translation from occurring,[364] they would promote a greater awareness of copyright on behalf of consumers and give authors more options without entirely restricting digital innovation. Indeed, many authors, keen on reaching new readers around the globe and gaining renown, may gladly acquiesce to the automatic translation of their content; the key is that their consent would be voluntary, not an imposition by a for-profit company. If the law required online translation service providers to use digital traffic lights, giving authors the ability to permit or reject translations, authors and service providers could work together: authors might actually be involved in translating their works and could receive compensation for them. The more in demand an author's works in different languages, the more the author could profit from partnering with an AI translation service provider. Just as Napster and Grokster paved the way for iTunes by revealing consumer demand for music online,[365] so could translation programs evolve in a direction that builds on people's desire to communicate without infringing copyright law.

We do not think metatags usage is the only viable approach, but the underlying idea has merit: by legally requiring software developers to give authors more choices concerning the automatic translation of their works online, both authors and service suppliers could profit from translations in

---

363. Ketzan, *supra* note 47, at 230–32.

364. *Id.* Ketzan notes that users could still paste text into a box, even if the translation program blocked the insertion of a particular URL. *Id.*

365. *See* Kelly Leong, *iTunes: Have They Created a System for International Copyright Enforcement?*, 13 NEW ENG. J. INT'L & COMP. L. 365 (2007).

the global marketplace.[366] Policymakers should establish new fair use guidelines for AI translations.

B. Second Conclusion: AI Translation Programs Should Promote Fair Use by Design[367]

In his book *Privacy's Blueprint: The Battle to Control the Design of New Technologies*, Woodrow Hartzog persuasively discussed the importance of product design in shaping behavior and protecting privacy: "Through signals, design helps define our relationships and our risk calculus when dealing with others. Design affects our expectations about how things work and the context within which we are acting."[368] In a similar vein, policymakers have suggested that privacy concerns should be addressed at the outset of product design, rather than post hoc once problems emerge.[369] In other words, "privacy initiatives should be 'proactive, not reactive, preventative, not remedial.'"[370]

Analogously, AI translation programs should use software design that facilitates compliance with fair use of copyrighted material through appropriate signaling, rather than encouraging copyright violations. Designers of these programs and their stakeholders "must begin with oversight and accountability in mind"[371] as this will encourage authors to participate in, rather than avoid, content-sharing platforms online. Users, meanwhile, will still be able to make *de minimis* use of translated copyrighted material without facing unreasonable hurdles.

---

366. *See* John S. Sieman, *Using the Implied License to Inject Common Sense into Digital Copyright*, 85 N.C. L. REV. 885, 887–89 (2007) ("First, an opt-in system preserves personal autonomy —the individual is not included unless he wants to be. This allows him the freedom to balance the costs and benefits of joining the system and make the choice of whether to join on his own. Second, opt-in systems place the burden on the system itself to gain individual participation. . . . First, "an opt-out system has the potential to be more valuable to the operator of the system than an opt-in system because it will include more people by default. Second, an opt-out system allows for the existence of systems that would otherwise be impracticable or prohibitively expensive. An opt-out system accomplishes this by placing the burden on individuals to remove themselves from the system.").

367. *See* Yanisky-Ravid & Hallisey, *supra* note 74.

368. WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES (2018).

369. KAMALA D. HARRIS & CAL. DEP'T OF JUST., PRIVACY ON THE GO: RECOMMENDATIONS FOR THE MOBILE ECOSYSTEM 1, 4 (2013), https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy /privacy_on_the_go.pdf [https://perma.cc/6CAK-44LG] ("Our recommendations, which in many places offer greater protection than afforded by existing law, are intended to encourage all players in the mobile marketplace to consider privacy implications *at the outset* of the design process.").

370. *See* Yanisky-Ravid & Hallisey, *supra* note 74, at 49.

371. *See* Knoll et al., *supra* note 76, at 640.

C. Third Conclusion: We Need WIPO and International Guidelines

The internet is intrinsically global in nature, AI translation tools are used worldwide, and international norms governing translations state that they are derivative works whose copyright lies with the original author. Therefore, to be effective, any solution to the issues raised must involve WIPO. International lawmakers must consult with one another, examining the ways in which different jurisdictions have handled conflicts between new web-based technology online and copyright law and working together to set new guidelines for AI translations.

This framework aims to support the public's right to communicate freely and engage with others on the web while protecting the rights of authors to profit from their original creative endeavors for a limited time. To the extent that we propose restrictions on the automatic translation of material, we do so also to emphasize the value of cultural diversity in all its nuances, which machines can never understand, let alone accurately replicate. Taken to extremes, copyright law can be a burdensome restriction on creativity and speech, but there is no need to seek an extreme solution here, and authors should not be forced to abandon all rights on the internet. Further, when we allow ourselves to be dazzled by AI translation technology, we not only ignore copyright, but we also lose sight of its inaccuracies and biases. In short, by paying a little more attention to these problems, lawmakers could readily solve them.

VII. Epilogue

The increased availability of translation services online has made it necessary for copyright law to address the unique issues such services raise concerning authors' moral rights and rights of reproduction and translation, balanced against the desirable goal of providing people worldwide with access to culture and ideas. Translation provides a useful linguistic tool in contexts that do not require perfect fluency, but its unrestrained use propagates biased language, violates authors' rights, and damages the market for high-quality translations (to the detriment of authors, translators and the general public). Therefore, the law should impose limits on the type and amount of material which for-profit companies may translate automatically online, amending the fair use doctrine to reference AI translations explicitly. It should also demand that AI translation software programmers implement fair use by design, in addition to privacy by design. The internet is global and so are languages; thus, WIPO should establish clear international guidelines concerning machine translation.

As we increasingly come to rely upon AI translation platforms, we should not lose sight of their weaknesses. Though certain readings of the

story of Babel may invite us to mourn the loss of a single language, others recognize that good translations celebrate global difference and depend upon the considerable linguistic dexterity, cultural fluency, and creative effort of translators.