The Business School for the World®

Working Paper

2023/24/EFE

AI-Driven Labor Substitution: Evidence from Google Translate and ChatGPT

Erdem Dogukan Yilmaz
Erasmus University Rotterdam, yilmaz@rsm.nl

Ivana Naumovska
INSEAD, ivana.naumovska@insead.edu

Vikas A.Aggarwal
INSEAD, vikas.aggarwal@insead.edu

Although artificial intelligence (AI) has the potential to significantly disrupt businesses across a range of industries, we have limited empirical evidence for its substitution effect on human labor. We use Google's introduction of neural network-based translation (GNNT) in 2016-2017 as a natural experiment to examine the substitution of human translators by AI in the context of a large online labor market. Using a difference-in-differences design, we show that the introduction of GNNT reduced the number of (human translation) transactions at both the overall market and individual translator levels. In addition, we show that GNNT had a stronger effect on translation tasks with analytical elements, as compared to those with cultural and emotional elements. In supplemental analyses, we document a similar pattern after the launch of ChatGPT using question and answer patterns in Stack Exchange forums. Our study thus offers robust and causal empirical evidence for a heterogeneous substitution effect of human tasks by skilled knowledge workers. We discuss the relevance of our findings for research on competitive advantage, technology adoption, and strategy microfoundations.

Key Words: Artificial Intelligence; Labor Substitution; Machine Translation

1. INTRODUCTION

Artificial intelligence (AI) has the potential to disrupt businesses across a range of industries by performing tasks traditionally undertaken by humans (Felten, Raj, & Seamans 2021; Felten, Raj, & Seamans, 2023). The release of AI technologies such as ChatGPT has sparked debate around the potential of AI to execute knowledge worker tasks. While the industry-wide effects of technologies such as this have yet to play out, scholars have suggested that AI is likely to enable some occupations, firms, and industries to flourish, while at the same time bringing about the destruction of others. AI is thus increasingly seen as a source of firm-level competitive advantage (Krakowski, Luger, & Rasich, 2022), with strategy scholars exploring the efficacy of AI in executing what were previously seen as exclusively human tasks (Puranam, 2018).

Despite the growing importance of AI, however, there is limited retrospective empirical evidence for the extent to which AI has substituted for human tasks (Felten, Raj, & Seamans, 2021; Toews, 2021; Acemoglu, Autor, Hazell, & Restrepo, 2022). Empirical work on AI's effects has generally been forward-looking, with scholars predicting that AI will substitute for human labor across a range of occupations, including highly skilled knowledge workers (Brynjolfsson, Mitchell, & Rock, 2018; Frank et al., 2019; Frey & Osborne, 2017). This research generally relies on expert opinions and forecasts (Felten et al., 2021; Felten et al., 2023; Frey & Osborne, 2017), and given that these are forward-looking opinion-based studies, their underlying assumptions may not be broadly accepted (Felten et al., 2021). Consequently, there remains lively debate as to which occupations and industries will ultimately be affected by AI, and to what extent (Felten et al., 2023; Brynjolfsson & Mitchell, 2017; Webb, 2019).

Past work has documented that pre-AI automation technologies have successfully substituted occupations that rely on routine tasks, such as basic accounting work, data entry and assembly line tasks (e.g., Autor et al., 2003; Acemoglu & Restrepo, 2020; Goldin & Katz, 1998; Mokyr, 1990). By contrast, AI technologies have the potential to substitute for more complex and non-routine tasks performed by knowledge workers (Brynjolfsson et al., 2018; Frank et al., 2019; Tong, Jia, Luo, & Fang, 2021).

To better understand the potential impact of AI on workers, occupations, firms, and industries, a critical starting point is to document the substitution effect of AI on tasks performed by skilled knowledge workers. Retrospective empirical evidence in this regard has been scant for several reasons. First, fine-grained data on AI adoption across firms is generally unavailable (Frank et al., 2019; Raj & Seamans, 2018). Second, the structure of tasks in many firms necessitates complementary technological investments (Brynjolfsson, Rock, & Syverson, 2021) as well as the deployment of specialized complementary human capital (Choudhury, Starr, & Agarwal, 2020). Third, the gradual adoption of AI over a longer timeframe makes it difficult to isolate the causal impact of AI on workers and firms (Brynjolfsson et al., 2021).

In this paper we provide casual empirical evidence for the substitution effect of AI. We respond to calls for the field of strategy to examine important business phenomena through a question-based, empirical lens without emphasizing formal hypothesis development (Graebner, Knott, Lieberman, & Mitchell, 2022). This approach is particularly useful in situations where "there is prevailing conventional wisdom but a dearth of robust research" (Graebner et al., 2022: 3). We employ a task-based level of analysis, an approach consonant with recent work on strategy microfoundations (Coff & Kryscynski, 2011; Barney & Felin, 2013; Greve, 2013; Felin, Foss, & Ployhart, 2015; Puranam, 2018). We exploit a natural experiment to identify the casual effect of AI on knowledge worker tasks, focusing on the substitutability of tasks *within* an occupation rather than on occupations as a whole, in response to calls for fine-grained empirical investigations into AI's effects (Frank et al., 2019).

Our industry context is language translation, in which AI-based machine translation (MT) has been increasingly employed (Jia, Carl, & Wang, 2019). Recently developed neural machine translation (NMT) methods have greatly advanced the state of the art in this industry (Jia et al., 2019; Yamada, 2019). Google introduced neural network (NMT)-based translation (GNNT) for different language pairs starting in November 2016. This introduction represented an exogenous shock to the translation occupation in that it suddenly and unexpectedly introduced a broadly and freely accessible AI alternative to human translations.

Translators are knowledge workers. Their work entails generating, editing, processing, and transforming knowledge and information. As such, translation relies on a high degree of what has been characterized as *non-routine cognitive tasks* (Pyöria, 2005). Whereas routine tasks can be performed by pre-AI automation technologies (Autor, 2014), AI has the potential to substitute for non-routine cognitive tasks. We focus on *non-routine cognitive tasks*, distinguishing between the *analytic* and *interactive* subtypes of such tasks (Autor, Levy, & Murnane, 2003; Spitz-Oener, 2006). While translations entail *non-routine cognitive tasks*, some are mostly *analytical* in nature, involving the interpretation of information in a different language (which we refer to as "regular translation"). On the other hand, some translation tasks are mostly *interactive*, requiring co-creation and adaptation of meaning to cultural and emotional elements, and are aimed at influencing emptions and behaviors (e.g., the translation of marketing and advertising material). Such work, relying on implicit knowledge as well as emotional and social intelligence of the translator, has been labeled "transcreation" by prior scholars (Pedersen, 2014).

We investigate the impact of GNNT on tasks involving transcreations versus regular translations using data from a major online labor market (OLM). OLMs, or so-called “outsourcing platforms” such as Amazon Mechanical Turk (MTurk), Upwork, Fiverr, Freelancer.com, Zooniverse, and Innocentive, match the flexible supply of labor with demand from across the globe (Agrawal, Lacetera, & Lyons, 2016). OLMs are particularly useful in assessing the impact of AI on human labor because they provide a digital, standardized, and complete archive of all transactions (Agrawal et al., 2016; Horton & Tambe, 2020). Our sample consists of 28,158 translation transactions conducted between January 2016 and May 2017, thus capturing a period of time both before and after the introduction of GNNT.

Using a difference-in-differences design we derive a rich set of results. First, we show that the introduction of GNNT reduced the number of (human translation) transactions at both the market and individual translator levels. We observe a significant drop in the number of language transactions for workers providing translation services for language pairs offered by GNNT.

Second, we find that regular translations (i.e., analytical tasks that entail interpreting information in a different language) experienced a drop of 13 to 20 percent (depending on the specification), whereas transcreations (i.e., interactive tasks that entail adaptation of meaning to cultural and emotional elements and influencing people) did not experience a drop.

Third, we observe a decline in the earnings of translation workers, amounting to $352,000 in the 6-month period after the launch of GNNT. We also investigate whether this drop in earnings was partially caused by a decrease in the prices charged by translators, rather than just a reduction in the number of transactions, but we find no evidence of price adjustments.

Fourth, we provide a set of supplemental analyses in which we utilize the launch of ChatGPT as a natural experiment, given its potential to disrupt and revolutionize various businesses and professions (Korinek, 2023). Specifically, we explore how the content of the forum Stack Overflow, a community for seeking code-related advice, has been affected by ChatGPT as compared to other forums (e.g., communities for seeking advice about travel, history, religion, philosophy). In line with our main GNNT-based findings, we document a more substantial effect of ChatGPT on the demand and supply for analytical content generated by human experts (e.g., coding), compared to less analytical content (e.g., history, philosophy, or religion).

In the remainder of this paper, we briefly discuss the background literature, before describing our empirical context, data, results, and supplemental analyses. Taken together, our results offer strong retrospective causal empirical evidence that AI can substitute for human tasks, with heterogeneity in the substitution effect as a function of the degree of cultural and emotional skills required by the task. We conclude with a discussion of implications for the strategy field.

2. BACKGROUND LITERATURE

Developments in AI have the potential to substantially affect firms, industries, and labor markets. Within the field of strategy, scholarly discourse has centered on how quickly and to what extent AI technologies will affect competition, strategy, and firms' human capital (Choudhury et al., 2020; Krakowski et al., 2022; Tschang & Almirall, 2021). The potential replacement by AI of tasks previously undertaken by humans raises critical questions that are at the heart of a microfoundational view on strategy (Barney & Felin, 2013; Felin, Foss, & Ployhart, 2015). Without a deeper understanding of the extent and nature of the human tasks at risk of being replaced by AI, however, it is difficult to fully appreciate the impact of AI on firms, industries, and markets. Furthermore, despite the potential significance of AI's impact, there is limited retrospective empirical evidence for its impact on human tasks.

Research on the impact of AI has generally been predictive, relying on theoretical models or expert opinions to forecast AI's substitution effect (e.g., Agrawal et al. 2019; Brzeski & Burk, 2015; Frey & Osborne, 2017; Pajarinen & Rouvinen, 2014; Webb, 2019). These forecasts vary in their estimated magnitude: 35 percent of the labor force being at risk of substitution in Finland (Pajarinen & Rouvinen, 2014), 47 percent in the US (Frey & Osborne, 2017), and 59 percent in Germany (Brzeski & Burk, 2015). A key limitation of such studies (e.g., as discussed in Arntz et al., 2017) is that their findings rely on experts' assessments of the substitution risk faced by particular occupations. Such opinions may be biased for at least two reasons. First, there is a lack of knowledge and consensus around the efficacy of AI technologies when engaging in different sets of tasks that have historically been performed by humans. As Brynjolfsson and Mitchell (2017, p. 1530) argue, there is "no widely shared agreement on the tasks where ML systems excel, and there is thus little agreement on the specific expected impacts on the workforce [...]." Second, when experts assess the substitution risk across occupations, they underestimate the *variety* of tasks within an occupation, and likely overestimate that risk (Arntz et al., 2017).

The popular narrative, however, suggests that the effects of AI could be significant for nearly half the US workforce, as well as the over 230 million knowledge worker roles worldwide (Daugherty & Wilson, 2019). Despite such fears, we lack a clear understanding of AI's capabilities and how it will affect knowledge workers, particularly with respect to the specific tasks it will impact. Additionally, there have increasingly been calls for more empirical research and theoretical advancement around the impact of AI on human tasks, particularly in the domain of non-routine cognitive tasks (Frank et al., 2019).

2.1. Conceptualizing and categorizing tasks by knowledge workers

Knowledge workers generally take on a range of tasks, some of which may be more susceptible to substitution threats than others. To understand the task-based substitution effect of technological advancements, it is important to conceptually differentiate tasks across two dimensions: (1) routine vs. non-routine, which captures the degree to which tasks involve explicitly repetitive procedures versus the need for dynamic intuition and adaptation capabilities (Acemoglu & Autor, 2011; Autor et al., 2003); and (2) manual vs. cognitive, which captures the degree to which physical versus intellectual skills are needed (Acemoglu & Autor, 2011; Autor et al., 2003). These two dimensions can jointly define four categories: a) routine manual, b) routine cognitive, c) non-routine manual, and d) non-routine cognitive.

Routine tasks (both manual and cognitive) follow explicit rules that can be specified in code and accomplished by machines. These tasks have already been replaced by pre-AI automation technologies (Autor, 2014). Routine manual tasks such as product assembly, painting, and welding have been automated with machinery (e.g., industrial robots), substituting for human labor in many situations (Acemoglu & Restrepo, 2020; Dixon, Hong, & Lu, 2021; Goldin & Katz, 1998; Mokyr, 1990). Routine cognitive tasks such as billing, accounting, and data entry have also been automated, with computers following pre-programmed instructions (Autor et al., 2003).

Non-routine tasks, on the other hand (both manual and cognitive), are more complex, involving skills such as problem-solving and intuition. These tasks entail adaptation to changing stimuli and conditions in the environment and cannot be pre-specified in computer code. Examples of non-routine manual tasks include driving a car through traffic and preparing a meal. Examples of non-routine cognitive tasks include medical diagnosis, legal advice, marketing campaign creation, and lab research. Until recently, such non-routine tasks seemed shielded from substitution (Autor, Levy & Murnane, 2003).

Scholars have suggested that AI will increasingly substitute for some non-routine tasks (Brynjolfsson et al., 2018; Frank et al., 2019), given the emerging evidence that AI technologies can in some cases outperform humans in tasks as diverse as medical diagnosis (Wang et al., 2019) and the maintenance of wind turbines (Franko, et al., 2020). However, the potential of AI may be more limited when non-routine cognitive tasks have interactive elements, such as the recognition of cultural norms or influencing people's emotions and behaviors, making them substitution-resistant, at least in the near future (Agrawal et al., 2019; Autor & Salomons, 2018; Marcus & Davis, 2019). Moreover, AI can also have a heterogeneous effect on workers in relation to their skill levels, given that more skilled knowledge workers may still have a stronger competitive advantage over AI technologies.

The central issues we investigate in this paper are two-fold. First, we seek to assess the substitution effect of AI. Our core question is whether and to what extent the introduction of AI technology (specifically, GNNT) resulted in the substitution of human tasks within the translation profession. Second, we seek to provide a more nuanced, task-based understanding of the potentially heterogeneous substitution effect across different types of non-routine cognitive tasks, thus examining non-routine cognitive (i) analytical versus (ii) interactive tasks.

3. INDUSTRY CONTEXT

Translation involves applying a "set of processes to render source language content into target language content in written form" (ISO 17100, 2015) in at least two working languages. These processes include translation, checking, revision, reviewing, and proofreading, and constitute numerous non-routine cognitive tasks. The market for translations has doubled in size over the last 10 years, reaching USD 49.6 billion in 2019, and is projected to increase further over the coming years (Launch, 2019). Despite the increase in demand for translations, the number of individuals employed in 2020 is significantly lower than in 2012 (Statista, 2020). Frey and Osborne (2017) estimate that the translation profession has a 38% chance of being automated in the next decade, placing it in the "medium risk" category of susceptibility to substitution. For reference, therapists and make-up artists face a probability of less than 1%. By contrast, professionals such as credit analysts and loan officers are more susceptible to substitution given AI's advantage over human counterparts in information processing, with an estimated substitution probability of 98% (Frey & Osborne, 2017). This suggests that the translation industry has heterogeneity in tasks that could help us identify which types are more or less susceptible to substitution by AI.

3.1. Google Neural Machine Translation (GNNT) as an industry shock

Early versions of machine translation had their roots in statistics-based translation algorithms developed by Warren Weaver and subsequently improved by researchers from IBM in the early-to-mid 1980s. Although this form of statistics-based machine translation saw broad adoption in the 2000s as a result of free services from Google, Microsoft, and Amazon, the quality of the translation offered was significantly inferior to that of human translation, and consequently did not offer a compelling alternative to human-based translation services (Wu et al., 2016).

A subsequent, and technologically superior, iteration of machine translation appeared in the form of "neural network machine translation." The idea of using neural networks to mimic human thinking has existed in scholarly discourse since the 1940s, although for decades the idea remained purely theoretical due to computational limits. With data processing and storage rapidly expanding over the past decade, however, the ability to employ neural networks in tasks across a broad range of industries increased substantially. These technological advances had a significant impact on a range of industries, such as computer vision and artificial intelligence in games (Moravčík. et al., 2017). Machine translation, likewise, adopted neural machine translation (NMT) models (Bahdanau, Cho, & Benigio, 2014; Luong & Manning, 2015). NMT considers the grammatical context of the sentences (as opposed to phrase-level, as in statistical machine translation), greatly contributing to legibility (Jia et al., 2019; Yamada, 2019). Prior to the entry of Google into this space, the distribution and reach of NMT for nonprofessional users was quite limited given its high price point (Jia et al., 2019).

Google introduced an NMT-based translation tool in 2016.[1] In contrast with prior offerings from other firms and independently operated human translators, the Google (GNNT) solution offered both high quality and broad reach (because of Google's distribution capabilities). Jia et al. (2019), for example, find that the fluency and accuracy of post-edited GNNT-generated translations is equivalent to that which is generated by humans from scratch. Similarly, Wu et al. (2016) find that GNNT reduces translation errors by 60% compared to Google's statistical machine translation for English-to-French and English-to-German pairs. After the November launch of eight language pairs, Google gradually replaced statistical machine translation with NMT for the remaining language pairs. By making the service freely available, Google was able to generate wide adoption. The subsequent growth of GNNT was significant: by 2016, Google Translate had surpassed one billion monthly active users, and was translating over 140 billion words each day (Schuster, Johnson, & Thorat, 2016).

The launch of GNNT was a shock to the translation industry, at its launch was unforeseen by the market, yet the product was immediately widely accessible. Discourse among the media, regulators, linguists, and translation workers focused on whether AI would bring an end to the

[1] Microsoft, like Google, developed and introduced NMT for 10 language pairs in November 2016: English with Arabic, Chinese Mandarin, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish. Seven of the pairs are identical to those introduced by GNNT: English with Mandarin Chinese, French, German, Portuguese, Japanese, Korean and Spanish. In contrast with Microsoft NMT, GNNT initially also covered the Turkish-English pair. In our empirical analysis, we keep the language pairs not covered by GNNT in the control group, making our findings more conservative because we would be underestimating the effect of GNNT if Microsoft had any significant effect on language pairs not affected by GNNT.

translation profession (Lewis-Kraus, 2016; Massey & Kiraly, 2019; Sung-won, 2017). Shortly after the launch of GNNT, the Secretary General of the International Interpretation & Translation Association, Kang Dae-young, stated, “Human translators and interpreters and those who seek to do these jobs in the future are increasingly facing concerns that they may lose their presence, as AI-based automatic translating technologies have rapidly been improved” (Sung-won, 2017).

Scholars raised similar concerns. In a 2019 exchange between Gary Massey and Don Kiraly, two leading translation scholars, Massey noted, “I have journalists coming up to me and asking: does neural machine translation (NMT) signal the end of the translation profession?” On that question, both Massey and Kiraly suggested that “there will be more work done by machine translation (MT)” but machines will not be able to perform good quality translations that involve “adaptive, creative, intuitive, ethically grounded work.” As Massey and Kiraly (2019) discussed, linguists have characterized such translations that entail “humanistic, interactive” components and seek to “influence opinions and decisions” as transcreations, distinguishing them from more analytic translations that rely on coded inputs.

The term “transcreation” combines two words, “translation” and “creation,” and captures the idea that translations need to consider the broader cultural context and nuance within which a text is situated. Transcreation can be described as “free of the literal,” and is used in market contexts to describe the translation of advertising material for different markets (Pedersen, 2014). The distinction between regular translation and transcreation is particularly useful for exploring the nuance within non-routine cognitive tasks, and in particular *non-routine cognitive and analytic* versus *non-routine cognitive and interactive* tasks, with the former entailing *regular translation* and the latter entailing *transcreation*.

For regular translations, such as news, essays, scientific papers, or legal and business documents, accuracy and precision are primary concerns. For these types of texts, AI has the potential to translate accurately, as the meaning is typically clear and unambiguous and translation process primarily focuses on the literal interpretation of the content (Viera, 2020). AI can recognize text patterns and translate the words used in everyday language or key technical terms, ensuring that the original text's meaning is preserved. By contrast, transcreation involves recreating a message in a different language, taking into account cultural and linguistic nuances, tone, and context. This is particularly crucial when translating marketing, advertising, poetry, or other creative content, where the message needs to resonate with the target audience and influence them. Transcreation requires not only translation skills but also creativity, as it involves adapting the original message to fit the cultural context of the target language (Pedersen, 2014).

AI may struggle with transcreation because it requires a level of creative and cultural understanding (Pedersen, 2014). This demands a level of emotional intelligence, which is acquired through personal and cultural experience, intuition, and personal skills. It is difficult to formalize or quantify such implicit knowledge, making it challenging for AI systems to learn and replicate. Furthermore, while there has been a substantial increase in the availability of parallel texts for translation, it is more challenging to obtain large amounts of transcreated content that would allow AI to implicitly capture (i.e., through unsupervised learning) cultural nuances, creativity, and context. Moreover, transcreation is a highly customized and subjective process, as different translators might produce distinct yet equally valid adaptations of the same content. AI systems, by their nature, tend to produce more consistent and standardized content, which may not be ideal for transcreation tasks that demand flexibility and creativity. Therefore, while AI may be able to produce regular translations of the content, it may not be able to capture the intended emotional or cultural nuances of the original message. We examine whether transcreations, which rest on *non-routine cognitive and interactive* tasks, are less likely to be substituted by NMT.

4. DATA AND METHODOLOGY

We use data from one of the largest online labor markets (OLMs). Human labor for translation is increasingly sourced through OLMs, which have become popular in recent years, with their jobs covering a wide variety of categories including translation, proofreading, writing, software development, product and logo design, administrative support, consulting, market research, customer service, and so on (Chen & Horton, 2016). OLMs are particularly useful in assessing the impact of AI on labor, since OLMs provide a digital, standardized, and complete archive of all their transactions (Agrawal et al., 2016; Chen & Horton, 2016). Our initial dataset included all freelancers that had at least one translation-related transaction in their portfolio between the years of 2001 and 2018. For the purpose of examining the effect of GNNT, we focused on the period between January 2016 and May 2017. We ended the sampling period in May because GNNT covered most of the other languages starting from June 2017.

To identify translation-related transactions, we examined the skills required by clients to complete the job using keywords (e.g., translation, interpretation, localization). We also relied on stemmed words in the job title and description—i.e., we reduced them to their root (e.g., the root forms of the words *"translation," "translated," "translate"* would be *"translat"*). In addition to relying on keywords that denote translations, we also examined words that indicated different languages and patterns in the data, as clients might refer to translation without using explicit keywords denoting translation (e.g., *"Articles from English to Italian,"* or *"Subtitles en > fr."*).

To identify the language that freelancers translated, we examined the freelancer's skill set. In some cases, clients would just write *"Translation"* without specifying the language pairs. In those cases, we looked at the freelancer's skill set to identify the language pair. If freelancers could translate from and into more than one language pair, and if we could not identify the language pair from the text of the transaction, we dropped those translation transaction from the data (26 transactions). Our final dataset consists of 28,158 transactions conducted between January 2016 and May 2017. Additional details about the sample construction can be found in the Appendix "Details of Sample Construction."

4.1 Interactive (i.e., transcreation) versus analytic (regular translations) translation tasks

Similar to other professions, translation tasks vary significantly across domains. We distinguish between analytic and interactive tasks, which correspond to two types of translations: *regular translations* (i.e., translations that are analytical in nature and could possibly be based on the literal and mechanical replacement of linguistic units) and *transcreations* (i.e., translation tasks that require "adaptive, creative, intuitive, ethically grounded work" [Massey & Wieder, 2019] such as the translation of advertising content or poetry).

To classify translation tasks into these two different categories, we first examined the tags used by clients in the job description. Apart from using tags such as *"translation"* and *"translation-english-spanish,"* clients on the online marketplace might also specify the content of the task as well as additional tasks required apart from the translation. For example, one of the clients on the online marketplace had written the following job description: *"I need singers from all around the world, who can translate and adopt lyrics to their own language and sing the translation"* and used "*translation," "voice-over," "singing,"* and *"audio-production"* tags. In another example, a client required the freelancer to translate a legal document, writing, *"I need someone to translate a court transcription from Russian to English,"* and used the *"legal-*

translation" tag. By using these tags, we first identified set of different transactions' categories. Specifically, we identified and grouped the following:

(1) Regular translations, consisting of general content (e.g., translation of journalistic texts, emails, letters, and product reviews) and technical content (e.g., medical documents with specialized text such as diagnosis and pharmaceutical observations, legal documents such as laws and privacy policies);[2]

(2) Transcreations, which consist of adapting and localizing a message from one language to another while maintaining its intent, style, tone, and goal. Such jobs rest on tasks of cultural and emotional adaptations required to make the content relevant to a new region and culture (e.g., adapting advertising materials and campaigns from one language to another); and

(3) Other types of translations which entail other types of services which are not core to the translation profession; audio/video services that require the translator to produce audio and video content, and certificate translations (e.g., diplomas, passports, marriage, and birth certificates) which require legal accreditation**.**

Following that, we hired two research assistants, randomly selected 4,000 observations, and instructed the assistants to categorize those translation transactions into these five groups: regular translations, technical content, transcreations, audio/video services, and certificate translations. They converged on 3,721 observations out of 4,000, and we eliminated the observations for which they were unable to reach the same classification. Next, we implemented preprocessing procedures by removing digits, non-English words, special characters, short words, punctuation, and URLs. Then we split the classified data into training and validation samples.

[2] Any translation task is considered to be a non-routine cognitive since each time translators translate a different text in a different context by using their non-routine cognitive skills such as critical thinking, judgement and decision making, and complex problem solving.

We used the BERT transformer model, a deep learning model used in natural language processing (NLP). Transformers are pre-trained on a massive dataset produced by prominent tech companies such as Google and Facebook, and are the current state-of-art in the NLP field (Acheampong, Nunoo-Mensah & Chen, 2021). These pre-trained models can be fine-tuned for specific classification purposes. The advantage of the BERT model is that it considers both the preceding and subsequent tokens at the same time and produces excellent classification results even with a small training dataset. Our models reached accuracy scores of 0.88. In the final step, we predicted the categories for translation transactions that have title and job descriptions but lack classification information.

[Insert Figures 1A & 1B about here]

To illustrate the classification, in Figures 1A and 1B we create word clouds for the raw data, and for each of the three categories: (1) regular translations (general and technical content), (2) transcreation, and (3) others (audio/video services and certificates). As Figure 1 shows, in the raw data, the most common words are related to translation. For the five sub-categories, the translation-related words are removed. Considering regular translations, we see that in the general content category clients did not provide any words that point to specific content, whereas in the technical content category the dominant words were “technical,” “legal,” and “medical.” In the transcreation category, the dominant word is “marketing.” Finally, in the category “Other,” which consists of tasks that entailed other complementary jobs in addition to translations, we see “audio,” “video,” “voice,” “certificate,” and “birth” as frequent words.

4.2. Identification strategy

We focus on events triggered by the launch of GNNT for different language pairs as exogenous shocks to the translation profession. In November 2016, eight different language pairs could be translated using GNNT.[3] In the period between March and April 2017, GNNT rolled out an additional 34 language pairs.[4] Importantly, GNNT represented a sudden and unexpected increase in the quality of machine translation. Therefore, our identification assumption for the causal interpretation is that the timing of the introduction of GNNT for different language pairs is unexpected by the market. This is a credible assumption, because while market participants might expect NMT to automatize some of the tasks in the translation profession, they cannot predict the exact timing of such a transition.

As noted above, prior to GNNT, Google Translate was based on statistical machine translation and the output from this tool was significantly inferior to that of human translators (Jia et al., 2019; Yamada, 2019). To isolate the impact of GNNT on the demand for translation of the different language pairs, we restricted our sample period to between January 2016 and May 2017, i.e., just before and after the introduction of the GNNT. We restricted our post-treatment period to May 2017 because, starting from June 2017, GNNT covered most of the language pairs in the data.

We conducted analyses both at the market level and at the individual freelancer level. Specifically, on the market level we examined how the overall volume of transactions was affected by the launch of GNNT, while at the freelancer level we examined how individual translators' transactions changed after the launch of GNNT. We used both Ordinary Least Squares (OLS) and Poisson to estimate the following regression specification using the language-pair and freelancer level panels:

$$Y_{it} = \alpha + \beta_1 * PostTreatment_{it} + \gamma_i + \delta_t + \epsilon_{it}$$

[3] From/to English -Mandarin, -French, -German, -Portuguese, -Japanese, -Korean, -Turkish and -Spanish.

[4] From/to English -Afrikaans, -Albanian, -Arabic, -Bulgarian, -Bengali, -Croatian, - Czech, -Danish, -Dutch, -Finnish, -Greek, -Gujarati, -Hebrew, -Hindi, -Hungarian, -Icelandic, -Indonesian, -Italian, -Japanese, -Kannada, -Malayalam, -Marathi, -Norwegian, -Polish, -Punjabi, -Romanian, -Russian, -Slovak, -Swedish, -Tamil, -Telugu, -Thai, -Ukrainian, -Vietnamese. In our data, we have observations for 26 of them.

where γi is language pair-fixed or freelancer-fixed effects and δt is time-fixed effects, for language pair or freelancer i and month of year t. Y_{it} indicates either the total number of transactions for a given language pair or the total number of transactions that a freelancer has in a given month. The term $PostTreatment_{it}$ equals one for a language pair or a freelancer that was affected by GNNT. Language pair-fixed effects control for time-invariant characteristics of language pairs such as the difficulty of translating one language from another or some languages being more context-dependent while others are more analytical. Time-fixed effects control for market-wide trends that affect each language pair equally.

This specification examines the difference between the number of transactions of language pairs or freelancers that were affected by GNNT, and that were not affected by GNNT, employing a difference-in-differences framework. If language pairs that are covered by GNNT experience a lower number of transactions, then we should observe a negative β_1 coefficient. All specifications cluster standard errors either at the language pair level or individual freelancer level to address the concern that transactions of the same language pairs or freelancers are likely to be correlated over time. Additionally, as recommended by Angrist and Pischke (2009), we do not include any control variables that might be affected by the treatment such as price, which could result in an inconsistent estimate of the treatment effect.

5. EMPIRICAL RESULTS

5.1. Summary statistics

Table 1 provides the summary statistics for the sample. Panel A provides summary statistics for key variables that vary at the language pair level. The main outcome variables are *Number of Transactions*, *Ln(Number of Transactions)*, *Ln(Dollar Amount)*, and *Ln(Price)*, measuring the number and dollar amount of transactions for each language pair. The main variable of interest is

PostTreatment, which is an indicator variable that is set to 1 if a language pair is affected by GNNT and zero otherwise. There are a total of 28,158 transactions across 47 language-pairs.

[Insert Table 1 about here]

5.2. Impact of GNNT on the number and dollar amount of transactions in the market

To begin, we plot the number of transactions that took place before and after the introduction of GNNT, comparing the cohort of treated language pairs with a control group of language pairs that were not treated. Figure 2 provides suggestive evidence that, before the introduction of GNNT, the treated and control language pairs displayed similar demand patterns. However, following the introduction of GNNT, treated language pairs experienced a decrease in the number of transactions compared to language pairs in the control group. Such a divergence occurs immediately after the introduction of GNNT and remains stable over time.

[Insert Figure 2 about here]

Table 2 presents estimates from this regression analysis on the number of transactions. The parameter estimate of *PostTreatment* is significant in both OLS (Column I) and Poisson (Column II) specifications, showing a decrease of 20.3 ($\beta = e^{-0.2283} – 1$, $p = 0.0007$) percentage points in the OLS specification and 12.5 ($\beta = e^{-0.1342} – 1$, $p = 0.0259$) in the Poisson specification.. Column III shows the OLS estimates on the total dollar amount of the transactions. In comparison to the control group, the total dollar amount of the transactions for the treated languages decreased by 31.9 ($\beta = e^{-0.3853} – 1$, $p = 0.0053$) percent. Considering the effect on the dollar amount, note that Clients on the OLM spent a total of $1,103,206 dollars for all treated language pairs after GNNT became available for those language pairs. This suggests that GNNT replaced $352,000 ((1,103,206 * (1+0.319)) - 1,103,206) worth of transactions that could otherwise have been completed by human translators.

In Figure 3, we plot the OLS specifications' time-varying estimates of the impact of GNNT on the number of transactions in the treated language pairs. In Figure 4, we plot the OLS specifications' time-varying estimates of the impact of GNNT on the dollar amount spent on the treated language pairs. As the figures indicate, the treated and control language pairs follow a similar trend before the treatment and after the treatment. However, the number of transactions and the total dollar amount of transactions of treated units tends to decrease. In addition, the effect grows in magnitude with time. This is because more and more people learn about the technology or realize its improved performance (see Google trends plot).

[Insert Table 2 about here & Figures 3 and 4 about here]

Next, we examine whether there was a heterogenous substitution effect for regular translations (entailing non-routine cognitive and analytic tasks) compared to transcreations (entailing non-routine cognitive and interactive tasks). Therefore, we examine the impact of GNNT on the number of transactions of regular translations versus transcreations. As noted above, the launch of GNNT, on average, had a negative impact on the demand for translations. In Table 3 we show the results for the differential effect on regular translations versus transcreations.[5] In the regular translation sample (columns I-III), the coefficients on *PostTreatment* are positive and significant for different dependent variables, whereas they are not statistically significant in the transcreation sample. Specifically, the number of transactions for regular translations dropped by 12.9 ($\beta = e^{-0.1386} - 1$, $p = 0.0192$) percent. Thus, we find support for H2. We also plot the time-varying effect on the number of transactions. As Figure 5 indicates, there is a declining trend in the number of regular translations, but there is no declining trend for the transcreation tasks. The drop in total dollar

[5] We report summary statistics for these analyses in Table A2.

amount in the regular translation sample is equal to 22.5 ($\beta = e^{-0.2550} - 1$, p = 0.0643) percentage points, totaling $164,000.

[Insert Table 3 about here & Figures 5 and 6 about here]

5.3. Impact of GNNT on number of transactions on individual freelancers

We now move our focus away from the translation market as a whole and explore the change in transactions for individual freelancers who were active in the treated languages prior to the introduction of GNNT. That is, we examine the change in transactions on the freelancer-month level. Our goal is similar to the previous analyses, with the difference that we compare the demand for freelancers with experience in treated languages to a counterfactual group of freelancers who were not specializing in those same treated languages. We restrict our analysis to those freelancers who were present on the OLM (i.e., had at least one transaction) one month before the introduction of GNNT. This is a conservative approach, because sellers with previous transactions are more likely to receive a job in a given time compared to those who lack previous transaction history (Benson, Sojourner, & Umyarov, 2020; Kokkodis & Ipeirotis, 2016).

The dependent variable corresponds to the number of transactions by each freelancer in a given month. We report summary statistics for the dataset on the freelancer-month level in Table 4. We estimate a regression similar to the baseline specification and report the results in Table 5—OLS specification in Column I, and a Poisson specification in Column II. The OLS specification shows a 4.7 ($\beta = e^{-0.0481} - 1$, p = 0.0906) percentage point decrease in a freelancer's number of transactions for the treated languages after the introduction of GNNT, whereas the Poisson specification shows a 9.1 ($\beta = e^{-0.0956} - 1$, p = 0.0573) percentage point decrease. Figure 7 depicts the time-varying estimates of the effect of GNNT on freelancers' transactions, using the OLS specification. These results at the freelancer level provide corroborating insights with the documented market-level outcomes.

[Insert Table 4 and 5 about here & Figures 7 about here]

5.4. Impact of GNNT on transaction price

Fourth, we examine whether the drop in the total dollar amount of the transactions was in part driven by a drop in the prices that freelancers charged, and not only the drop in the number of transactions. Therefore, we examine the effect that the launch of GNNT had on the average price of the transactions. Specifically, the unit of analysis is language pair per month, and the dependent variable is the average dollar amount paid for all of the transactions in that given month. We find a negative effect on the change in the average prices, but that effect is not statistically significant. We also plot the results for price analyses in Figure 11. Although the figure indicates an overall declining trend for the average price of transactions, it is not statistically distinguishable from zero.

This suggests that the drop in the dollar amount observed in Table 7 is driven primarily by the drop in the number of transactions rather than a drop in the average amount paid per transaction. While one can expect that GNNT as a competitor can trigger a drop in the value of human translations, our findings may simply imply that the freelancers' price response was delayed. Indeed, freelancers on OLMs typically do not react to demand fluctuations with changing prices, but rather with performing more or fewer tasks—i.e., changing the supply (Cullen & Farronato, 2021). However, in a longer time frame, they might adjust the prices along with supply.

[Insert Table 6 about here & Figure 8 about here]

5.5. Impact of GNNT on search trends as a proxy for market demand

We also sought to examine the representativeness of our findings for the broader labor market of translators, given that our sample comes from a specific OLM (Roth, 2018). One way to assess representativeness—at least with respect to demand for machine translation—is to observe the search engine query volumes for different language pairs related to Google Translate. In this exercise, therefore, we examine how the search query volumes changed for the affected languages. Since Google is a widely used search engine throughout the world, it might be one of the best sources of information about the "real world" labor market (Baker & Fradkin, 2017). In this section, therefore, we replicate our analysis with data obtained from Google Trends. Specifically, for each treated and control group that appears in our previous data set, we collect search trend data from Google by using Google Translate search terms specific to the relevant language pairs. For instance, for the English-Spanish language pair, we use the search terms such as "Google Translate English Spanish." Our observation period is the same: it starts in January 2016 and ends in May 2017. If we observe a positive coefficient on the *PostTreatment* variable, this will indicate that the demand for machine translation increases when the AI-based tool GNNT is introduced to the market. Consistent with our expectation, search query volumes for affected languages increase after the treatment by 8.9 ($\beta = e^{0.0856} - 1$, $p = 0.0185$) percentage points, as reported on Table 6. Similarly, we plot the time-varying impact of GNNT against the demand for Google Translate on Figure 10. In line with our results, the former head of Google Brain Team stated in one of his presentations at an invited talk that the use of the Korean-English language pair in Android devices has increased by 75% between June 2017 to December 2017.[6]

[Insert Table 7 about here & Figure 8 about here]

5.6. Evidence with coarsened exact matching (CEM)

As a supplement, we use the coarsened exact matching method, following Iacus et al., (2012). The advantage of this methodology is that the characteristics of the treated units before the policy change can be better approximated by a combination of untreated units compared to an unweighted

[6] https://workshop2017.iwslt.org/downloads/InvitedTalk1-Slide.pdf

group of untreated units (Iacus, King, & Porro, 2012). Accordingly, we construct a control sample for the first treatment cohort by using a weighted combination of potential control language pairs. In constructing control groups, we use pretreatment values of language pairs' average prices per transactions, market size (i.e., the total value of transactions in a given month), and the trend in their number of transactions as matching covariates. As Table 4 shows, both the number of both transactions of the treated language pairs and the dollar amount of transactions substantially decreased after the introduction of GNNT. The results are consistent with the baseline results. These results provide strong support for the substitution of human translation by GNNT.

[Insert Table 8 about here]

5.7. Using Alternative Treatment Data and Accounting for Staggered Adoption

We performed additional checks to test the robustness of our results. This included using an alternative treatment starting date for the Chinese-English pair, and we only included observations up to the point at which the second and third cohorts of treatment remained in the control group, to avoid staggered treatment timing which recent econometrics literature has highlighted as problematic (see the discussion by Baker, Larcker & Wang, (2022)). This approach resulted in having only one treatment cohort, similar to the case of classical difference-in-differences analyses. These results are reported in the appendix. The results were consistent across the different specifications.

5.8. Empirical extension: Evidence from ChatGPT

While translation tasks can be considered examples of prediction tasks, AI can also be employed for generative tasks, such as creating code or content. One notable example of generative AI technology is ChatGPT, introduced by OpenAI on November 30th. Having attracted significant attention, ChatGPT, with its ability to produce high-quality written text and widespread reach, holds the potential to disrupt and revolutionize various businesses and professions. ChatGPT boasts advanced capabilities in ideation, writing, background research, coding, data analysis, and mathematical derivations (Korinek, 2023).

The introduction of ChatGPT has also sparked considerable interest among a wide range of professionals, including academics, marketing experts, software developers, and engineers. However, despite its popularity, concerns have been raised about the accuracy and quality of ChatGPT's output. Content can sometimes be erroneous and misleading (Glorioso, 2023). For example, Stack Overflow has banned the use of ChatGPT for its tendency to generate answers that appear correct but are actually incorrect.[7] Additionally, language professionals have observed that ChatGPT's translations may lack accuracy and detail. In fact, ChatGPT has achieved noticeably inferior results in all translation-related evaluations when compared to both Google Translate and DeepL (Jiao, Wang, Huang, Xing & Tu, 2023). Experts claim that "ChatGPT is unlikely to beat machine translation engines in terms of accuracy simply because MTs are trained with domain-specific data and terminology, and ChatGPT is not" (Wordspath Team, 2023). These factors highlight ChatGPT's limitations when it comes to analytical tasks.

In addition, the abilities of ChatGPT to engage in interactive tasks, which involve emotional intelligence, creativity, and cultural understanding of the audience, are also limited. For instance, poems written by ChatGPT are filled with overused expressions and cliché rhymes and have no artistic value (Hunter, 2023). Moreover, algorithms such as ChatGPT may be restricted in the extent to which they can in fact understand the "meaning of their responses" (Bender & Koller, 2020), which could hinder their ability to appropriately address cultural nuances.

[7] See https://meta.stackoverflow.com/questions/421831/temporary-policy-chatgpt-is-banned

If our findings about the ability of AI to substitute for analytical tasks can be generalized, we should observe a more substantial effect of ChatGPT on the creation of analytical content, like coding, compared to less analytical content, such as discussions of philosophy, religion, building imaginary worlds, or content that may require up-to-date or personal knowledge of other topics like travel and science fiction. To examine the substitution effect in a different setting with generative AI technology, we collected data from various Stack Exchange forums, including Stack Overflow (a community for code-related help), Christianity (christianity.stackexchange.com), cryptocurrency (crypto.stackexchange.com), history (history.stackexchange.com), Islam (islam.stackexchange.com), philosophy (philosophy.stackexchange.com), science fiction (scifi.stackexchange.com), travel (travel.stackexchange.com), and imaginary worldbuilding (worldbuilding.stackexchange.com).

In our analysis, we examine how the content on Stack Overflow has been affected by ChatGPT compared to other forums in terms of the quantity of questions and answers posted. We use data at the week-forum level and compare the number of questions and answers posted before and after the introduction of ChatGPT. Similar to previous analyses, we adopt a difference-in-differences design.[8] We report the results from analyses in Table 9. In Log-OLS specifications, we find a 12.1 ($\beta = e^{-0.1295} - 1$, $p = 0.0307$) percentage point decrease in the number of questions asked on Stack Overflow compared to other question and answer forums, whereas although it seems like there is a decrease in the number of answers posted, the effect in Log-OLS specification is marginally significant ($p < 0.15$). In Poisson specification, this corresponds to a 16.1 ($\beta = e^{-0.1763} - 1$, $p = 0.0000$) percentage points and to a 18.5 ($\beta = e^{-0.2054} - 1$, $p = 0.0004$) percentage points

[8] Given that there is single treatment date, we use a classical difference in differences as opposed, two ways fixed effect.

decrease, for questions and answers respectively. We plot the average number of questions asked and answers posted in Figure 10. While the number of questions asked seems to increase after a significant drop, it does not catch up with the questions asked in the post-treatment period. The number of answers posted, on the other hand, seems to maintain its declining trend. We plot the average number of questions asked and answers posted in Figure 10.

As ChatGPT was launched before the Christmas period, during which activity on Stack Overflow may decline more compared to other forums, we created an alternative sample by including the top programming languages that ChatGPT is most proficient in. To do this, we asked ChatGPT which programming language it is most proficient in a hundred times, and only included those languages that consistently appeared in the sample. These languages are Python, JavaScript, Java, C++, C#, and PHP. We then compared the number of questions and answers posted in the year before ChatGPT's launch to those in the year of its launch. We report the results from this exercise on Table 10 and plot the average numbers of questions and answers on figure 11.

These analyses suggest that users may be relying less on Stack Overflow for assistance, leading to a reduced need to ask questions on Stack Overflow. While one might expect the number of questions asked on Stack Overflow to decrease given ChatGPT can potentially substitute the answers that would be posted by other users, the results on the number of answers are more surprising. Users may use ChatGPT to increase their productivity and the number of replies they post. There are several alternative explanations for this finding: First, users might be asking more difficult questions, making it more challenging for others to answer. Second, when there is a competing platform or solution available in the market, they may be demotivated to further contribute to that community.

These results are in line with the results that we obtained from GNNT exercise. In particular, in both settings, we found that the substitution effect of AI on human labor is heterogeneous and depends on the nature of the tasks involved. AI may be more effective at substituting human labor in analytical tasks, while its impact may be limited when it comes to tasks requiring emotional intelligence or contextual understanding.

[Insert Tables 9 and 10 about here & Figures 10 and 11 about here]

6. DISCUSSION

This study examines the impact of AI on the demand for human labor within the context of language translation by using the introduction of Google neural network-based translation (GNNT) as a shock to the industry. We find evidence that GNNT substitutes for human labor, and that this negative effect is more pronounced for translations that rely mostly on analytical tasks. Tasks that require interactive, adaptive, creative, and intuitive work were not affected by the launch of GNNT.

6.1. Implications for work on strategic human capital

Our study contributes to the debate on AI's potential to replace human labor (e.g., Autor, 2014; Choudhury et al., 2020; Frank et al., 2019; Raisch & Krakowski, 2021) by providing causal evidence of the task-level substitution effect. Our findings support the notion, moreover, that AI replaces different tasks within an occupation to different extents (Acemoglu & Autor, 2011; Autor et al., 2003). Second, our findings on the heterogeneous effect of AI on different tasks contribute to the innovation literature on strategic human capital. While there is no explicit labor division between AI and human labor (Puranam, 2021), we show that, at least initially, AI is more likely to replace analytic tasks within an occupation, while tasks that require the processing of cultural and emotional elements are likely to remain somewhat resistant to substitution resistant.

Therefore, workers would need to devote more effort to developing skills where they have a competitive advantage over AI. For instance, they might put greater effort into tasks that involve interactions with people, tacit knowledge, social skills, and creativity (Agrawal, Gans, & Goldfarb, 2018; Choudhury et al., 2020). Indeed, according to responses to one of the recent surveys conducted with language translation professionals, this process may already be underway: “As technical translation becomes increasingly automated, technical translators move into marketing translation” (EC, CIOL, and ITI, 2017, p. 23). Such a substitution effect implies that firms should carefully examine adoption in relation to their workforce’s skill set and their own organizational context (Agrawal, Gans, & Goldfarb, 2021). At the labor market level, our findings do not suggest that AI will displace the translation profession, but that it will likely lead to the restructuring of the task content of the profession, as workers adapt to offer services that are AI-resistant.

We know that new corporate practices and technologies are often met with skepticism by workers (Bailey & Barley, 2011; Lapointe & Rivard, 2007; Pachidi, Berends, Faraj, & Huysman, 2021). Our findings point to the need for workers and firms to anticipate the risks stemming from AI-based substitution, and to adapt their repertoire of skills. Such a process can also be facilitated through appropriate incentive schemes (Gambardella, Panico, & Valentini, 2015). For example, post-editing of machine translation (PE) has been established as a translation service, as certified by the International Standards Organization (ISO 18587, 2017), and one would expect more workers to specialize in this domain of the translation profession. While AI has the potential to disrupt established work dynamics, we have yet to understand its heterogeneous effects. Thus, workers should monitor the advancement of new neural network-based tools to better anticipate the threats and opportunities stemming from AI.

6.2. Implications for work on resource allocation

As AI continues to advance, it is likely to have a significant impact on how firms allocate their resources. Prior work shows that novel technologies may impact the allocation of resources and division of labor (Leonardi & Barley, 2010; Orlikowski & Scott, 2008) by means of automating repetitive tasks and thereby freeing up certain human resources (Autor et al., 2003). Similarly, AI is likely to affect resource allocation, albeit through the substitution (or augmentation) of non-routine tasks. Specifically, as our study shows, AI systems have the capacity to automate and thus substitute for some non-routine analytical tasks that were previously executed by humans, thereby releasing resources that can be repurposed elsewhere within a firm. For instance, AI systems can process vast amounts of data and identify patterns that would be insurmountable for humans to discern, enabling firms to direct knowledge workers who were formerly engaged in analytical tasks towards interactive tasks.

The reallocation of tasks resulting from the advancement of AI implies that its substitution effect carries important implications for firms' competitive advantage. Notably, the tasks executed by knowledge workers constitute a crucial resource that confers a competitive advantage upon firms (Spender & Grant, 1996). Knowledge workers must amass extensive training and experience to cultivate their domain-specific cognitive capabilities (Helfat & Peteraf, 2015; Kunc & Morecroft, 2010). Considering the substantial investment required for the development of such skills, the pool of available candidates remains limited. For example, traditionally, generating insights from data demanded significant cognitive abilities for data collection, processing, and analysis. However, with AI able to execute many of the tasks previously executed by knowledge workers, numerous firms employing these workers face the prospect of losing their competitive advantage as the cognitive capabilities they possess become available at scale in the market.

To acquire a competitive advantage through AI, firms that rely on non-routine cognitive analytical tasks will not just need to invest in new technologies and systems, they will also need to train their workers to use these technologies effectively. Knowledge workers must learn how to manage and interpret the output produced by AI (Waardenburg, Huysman, & Sergeeva, 2022). For example, while AI-based systems are very good at identifying patterns in medical data, trained healthcare professionals still need to interpret the output to effectively use the insights generated by these systems. In some cases, firms may even need to hire "algorithmic brokers" (Kellogg, Valentine, & Christin, 2020) who simplify the usage of AI systems by interpreting output for users (Henke, Levine, & McInerney, 2018).

6.3. Implications for work on technology diffusion and adoption

Despite the potentially significant outcomes of AI, we lack an understanding of firm-level factors that can facilitate the adoption and diffusion of AI technologies. Our findings can shed light on such questions and inform research on their diffusion. If we conceptualize firms as bundles of tasks (Garicano & Wu, 2012), then our findings suggest that firms whose knowledge workers perform a higher portion of non-routine cognitive analytical tasks—rather than non-routine cognitive interactive ones—will be more likely to adopt AI technologies.

Indeed, tasks that employees conduct at their jobs influence firms' adoption of new technologies. We already know that firms whose workers perform highly repetitive or routine tasks were more likely to adopt pre-AI automation technologies to increase efficiency (Autor et al., 2003). On the other hand, there is emerging evidence that AI can carry out more complicated, non-repetitive tasks that, until recently, could only be completed by humans (Acemoglu et al., 2022).

Our analyses demonstrate that, while AI can effectively perform analytical cognitive non-routine tasks, it does not replace interactive tasks. This implies that firms whose employees engage in non-routine cognitive tasks may be more inclined to adopt AI technologies. Some industries and firms in those industries are more reliant on analytical cognitive non-routine tasks, such as software publishers, research, and computer system designs (Autor et al., 2003). Such firms may therefore be particularly inclined to use AI to augment or replace some of their human workforce. For example, a firm in software publishing could use AI for automated software testing and documentation, as well as customer support. While the use of AI may substitute for some of the tasks performed by engineers, such professionals can also rely on AI to improve their performance related to analytical tasks (e.g., identifying bugs), pointing to the augmenting potential of AI technologies.

By contrast, interactive non-routine cognitive tasks involve the application of complex problem-solving and decision-making abilities, as well as an understanding of cultural and emotional context, and often involve interaction with and persuasion of individuals. Examples of interactive cognitive non-routine tasks include negotiating, persuasion, and teaching, and as our results suggest, firms whose workers perform interactive non-routine cognitive tasks may be less likely to adopt AI technologies. Instead, firms may rely on the expertise and judgment of human workers to perform these tasks. However, it is important to note that AI technologies may still be able to augment the performance of interactive non-routine cognitive tasks by providing relevant information and decision-making support to human workers.

6.4. Implications for work on strategy microfoundations

Our task-based view of the impact of AI has theoretical implications for the ongoing conversation around the microfoundations of strategy (Barney & Felin, 2013; Felin, Goss, & Ployhart, 2015; Foss & Pedersen, 2014). At the heart of this discussion is the idea that understanding the locus of firm-level competitive advantage necessitates a focus on the organizational processes that aggregate beliefs, knowledge, and decisions from the individual to the organization-level of analysis (Aggarwal, Posen & Workiewicz, 2017; Barney & Felin, 2013; Puranam, 2018). The processes by which individuals and groups of individuals interact can help explain how and why firms are able to mobilize resources to address new and emerging competitive threats, as well as engage in ongoing innovative activity (Davis & Aggarwal, 2020). The retrospective empirical evidence we provide in this study for the types of tasks and workers that are higher risk of being substituted for by AI thus provides important evidence-based insights into the sets of innovation processes that may need to be redesigned as AI adoption continues.

Recent work offers various theoretical suggestions as to how human-AI collaboration may be most fruitful (Puranam, 2021). By unbundling the set of tasks pursued by knowledge workers, our work provides empirical evidence that can guide future theoretical work in this spirit by identifying the sets of organizational interfaces within firms that are likely at risk of change due to AI. Employees whose key activities include non-routine cognitive tasks are likely to employ a portfolio of skills in their jobs, some of which may be more susceptible to substitution by AI than others. As a result, these individuals may need to alter the ways in which they receive, process, and transmit information to others within the firm (Joseph & Gaba, 2020). The changing nature of the skills these individuals are able to contribute to the organization has implications not only for formal organizational structure, but also for the value of individuals' informal networks (Haas & Hansen, 2007; Hansen, Mors, & Lovas, 2005). In sum, AI is likely to have a heterogeneous impact not only across individuals within firms, but also for the particular skills individuals bring to bear in order to create value for the firms in which they operate (Coff & Kryscynksi 2011; Felin & Hesterly, 2007). Understanding the ways in which AI will shape firm-level competitive advantage thus necessitates a disaggregated, task-based view of the impact of AI, of which we offer an empirical first step.

6.5. Limitations

We highlight several limitations of this paper. First, we study a market that has been affected by one of the most advanced AI technologies, NMT. Indeed, in recent years there have been many significant advances in language-related AI and NMT, which in part has been driven by the availability of training data. However, in other settings such as speech recognition, vision, and surgery, AI may be further away from being a competent and reliable substitute for human workers. Future research may examine how technological advancement (including the quality of training data) and sociological dynamics interact with and shape the substitution effect of AI on human labor.

Second, and related to the first point raised above, our research is limited to examining the substitution effect of AI on the execution of a prediction task, namely language translation, which is at a higher risk of being substituted for by AI than judgment tasks (Agrawal et al., 2019). However, as the frequency and capability of AI practices spread across industries, not only prediction tasks but also some decision-making tasks may be at risk of being substituted for by AI. Thus, it is important for future research to examine the impact of AI on such tasks.

Third, our data only run through May 2017, and it is likely that GNNT has already advanced through the years, necessitating an update to our study's conclusions, perhaps through a different identification strategy. For instance, with increasing computational capabilities and the availability of training data, the substitution effect of GNNT might already be larger than what we report in our analyses.

Fourth, in our data, some of the translation transactions were missing a job description, and the job title did not contain relevant information about the type of the translation task. For those translation transactions, we assumed that translation content was general, as clients are more likely to disclose special requirements if they exist. Future research might collect more detailed data on translation projects, including the text of the translation, to see if our findings hold. These limitations notwithstanding, we believe that our study provides credible empirical evidence pointing to the substitution effect of AI on human labor.

REFERENCES

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics* (Vol. 4: 1043-1171). Elsevier.

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, *40*(S1), S293-S340.

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In *The economics of artificial intelligence: An agenda* (pp. 197-236). Chicago, IL: University of Chicago Press.

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, *33*(2), 3-30.

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, *128*(6), 2188-2244.

Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, *54*(8), 5789-5829.

Aghion, P., Antonin, C., Bunel, S., & Jaravel, X. (2022). The effects of automation on labor demand: A survey of the recent literature. *Available at SSRN DP16868.*

Aggarwal, V. A., Posen, H. E., & Workiewicz, M. (2017). Adaptive capacity to technological change: A microfoundational approach. *Strategic Management Journal*, *38*(6), 1212-1231.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Boston, MA: Harvard Business Press.

Agrawal, A., Gans, J., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy*, *47*, 1-6.

Agrawal, A., Gans, J., & Goldfarb, A. (2021). AI Adoption and system-wide change. NBER working paper 28811, National Bureau of Economic Research.

Agrawal, A., Lacetera, N., & Lyons, E. (2016). Does standardized information in online markets disproportionately benefit job applicants from less developed countries?. *Journal of International Economics*, *103*, 1-12.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.Arntz, M., Gregory, T., & Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters*, *159*, 157-160.

Autor, D. (2014). Polanyi's paradox and the shape of employment growth. NBER working paper 20485, National Bureau of Economic Research.

Autor, D. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3-30.

Autor, D., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279-1333.

Autor, D., & Salomons, A. (2018). Is automation labor-displacing? Productivity growth, employment, and the labor share. NBER working paper 24871, National Bureau of Economic Research.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473.*

Bailey, D. E., & Barley, S. R. (2011). Teaching-learning ecologies: Mapping the environment to structure through action. *Organization Science*, *22*(1), 262-285.

Baker, S. R., & Fradkin, A. (2017). The impact of unemployment insurance on job search: Evidence from Google search data. *Review of Economics and Statistics*, *99*(5), 756-768.

Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates?. *Journal of Financial Economics*, *144*(2), 370-395.

Barney, J. A. Y., & Felin, T. (2013). What are microfoundations?. *Academy of Management Perspectives*, *27*(2), 138-155.

Bender, E. M., & Koller, A. (2020, July). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 5185-5198).

Benson, A., Sojourner, A., & Umyarov, A. (2020). Can reputation discipline the gig economy? Experimental evidence from an online labor market. *Management Science*, *66*(5), 1802-1825.
Bessen, J. (2018). AI and Jobs: The role of demand. NBER working paper 24235, National Bureau of Economic Research.
Billmeier, A., & Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, *95*(3), 983-1001.
Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530-1534.
Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy?. *AEA Papers and Proceedings*, *108*, 43-47.
Brynjolfsson, E., Hui, X., & Liu, M. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science, 65*(12), 5449-5460.
Brynjolfsson, E., Frank, M. R., Mitchell, T., Rahwan, I., & Rock, D. (2020). Quantifying the impact of machine learning on work. Working paper.
Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, *13*(1), 333-72.
Brzeski, C., & Burk, I. (2015). Die Roboter kommen. Folgen der Automatisierung für den deutschen Arbeitsmarkt. *INGDiBa Economic Research*, *30*.
Chen, D. L., & Horton, J. J. (2016). Research note—Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts. *Information Systems Research*, *27*(2), 403-423.
Chiacchio, F., Petropoulos, G., & Pichler, D. (2018). The impact of industrial robots on EU employment and wages: A local labour market approach. Bruegel working paper No. 2018/02.
Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, *41*(8), 1381-1411.
Coff, R., & Kryscynski, D. (2011). Invited editorial: Drilling for micro-foundations of human capital–based competitive advantages. *Journal of management*, *37*(5), 1429-1443.
Cronin, M. (2013). *Translation and globalization*. London: Routledge.
Cullen, Z., & Farronato, C. (2021). Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Management Science*, *67*(7), 3985-4003.
Dauth, W., Findeisen, S., Suedekum, J., & Woessner, N. (2021). The adjustment of labor markets to robots. *Journal of the European Economic Association*, *19*(6), 3104-3153.
Davis, J. P., & Aggarwal, V. A. (2020). Knowledge mobilization in the face of imitation: Microfoundations of knowledge aggregation and firm-level innovation. *Strategic Management Journal*, *41*(11), 1983-2014.
Dixon, J., Hong, B., & Wu, L. (2021). The robot revolution: Managerial and employment consequences for firms. *Management Science*, *67*(9), 5586-5605.
EC, CIOL, ITI (2017). *2016 UK translator survey: Final report. European Commission.* Retrieved from https://ec.europa.eu/unitedkingdom/education/languages/2016- survey-of-translators-based-in-UK_en.
Felin, T., & Hesterly, W. S. (2007). The knowledge-based view, nested heterogeneity, and new value creation: Philosophical considerations on the locus of knowledge. *Academy of management review*, *32*(1), 195-218.
Felin, T., Foss, N. J., & Ployhart, R. E. (2015). The microfoundations movement in strategy and organization theory. *Academy of Management Annals*, *9*(1), 575-632.
Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, *42*(12), 2195-2217.
Felten, E., Raj, M., & Seamans, R. (2023). How will Language Modelers like ChatGPT Affect Occupations and Industries?. *arXiv preprint arXiv:2303.01157*.

Foss, N. J., Pedersen, T., Reinholt Fosgaard, M., & Stea, D. (2015). Why complementary HRM practices impact performance: The case of rewards, job design, and work climate in a knowledge-sharing context. *Human Resource Management*, *54*(6), 955-976.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences, 116*(14), 6531-6539.

Franko, J., Du, S., Kallweit, S., Duelberg, E., & Engemann, H. (2020). Design of a multi-robot system for wind turbine maintenance. *Energies*, *13*(10), 2552.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological Forecasting and Social Change*, *114*, 254-280.

Gambardella, A., Panico, C., & Valentini, G. (2015). Strategic incentives to human capital. *Strategic Management Journal*, *36*(1), 37-52.

Garicano, L., & Wu, Y. (2012). Knowledge, communication, and organizational capabilities. *Organization science*, *23*(5), 1382-1397.

Glen, S. (2022). ChatGPT writes code, but won't replace developers. TechTarget. Retrieved from https://www.techtarget.com/searchsoftwarequality/news/252528379/ChatGPT-writes-code-but-wont-replace-developers

Glorioso, C. (2023, February 23). Fake News: ChatGPT Has a Knack for Making Up Phony Anonymous Sources. *NBC New York*. Retrieved from https://www.nbcnewyork.com/investigations/fake-news-chatgpt-has-a-knack-for-making-up-phony-anonymous-sources/4120307/

Goldin, C., & Katz, L. F. (1998). The origins of technology-skill complementarity. *The Quarterly Journal of Economics*, *113*(3), 693-732.

Haas, M. R., & Hansen, M. T. (2007). Different knowledge, different benefits: Toward a productivity perspective on knowledge sharing in organizations. *Strategic management journal*, *28*(11), 1133-1153.

Hansen, M. T., Mors, M. L., & Løvås, B. (2005). Knowledge sharing in organizations: Multiple networks, multiple phases. *Academy of Management journal*, *48*(5), 776-793.

Helfat, C. E., & Peteraf, M. A. (2015). Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic management journal*, *36*(6), 831-850.

Henke, N., Levine, J., & McInerney, P. (2018). Analytics translator: The new must-have role. *Harvard Business Review*.

Horton, J. J., & Tambe, P. (2020). The death of a technical skill. *Unpublished Manuscript*.

Hunter, W. (2023, February 13). What Poets Know That ChatGPT Doesn't. The Atlantic. Retrieved from https://www.theatlantic.com/books/archive/2023/02/chatgpt-ai-technology-writing-poetry/673035/

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, *20*(1), 1-24.

ISO (2015). *ISO 17100:2015 Translation services — Requirements for translation services.* Retrieved from https://www.iso.org/standard/59149.html.

ISO (2017). *ISO 18587:2017 Translation services — Post-editing of machine translation output — Requirements.* Retrieved from https://www.iso.org/standard/62970.html.

Jia, Y., Carl, M., & Wang, X. (2019). Post-editing neural machine translation versus phrase-based machine translation for English–Chinese. *Machine Translation*, *33*(1), 9-29.

Joseph, J., & Gaba, V. (2020). Organizational structure, information processing, and decision-making: A retrospective and road map for research. *Academy of Management Annals*, *14*(1), 267-302.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366-410.

Kokkodis, M., & Ipeirotis, P. G. (2016). Reputation transferability in online labor markets. *Management Science*, *62*(6), 1687-1706.

Korinek, A. (2023). *Language Models and Cognitive Automation for Economic Research* (No. w30957). National Bureau of Economic Research.

Krakowski, S., Luger, J., & Raisch, S. (2022). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*.

Kunc, M. H., & Morecroft, J. D. (2010). Managerial decision making and firm performance under a resource-based paradigm. *Strategic management journal*, *31*(11), 1164-1182.

Lapointe, L., & Rivard, S. (2007). A triple take on information system implementation. *Organization Science*, *18*(1), 89-107.

Leonardi, P. M., & Barley, S. R. (2010). What's under construction here? Social action, materiality, and power in constructivist studies of technology and organizing. *The Academy of Management Annals*, *4*(1), 1-51.

Lewis-Kraus, G. (2016, December 14) The great A.I. awakening. *The New York Times*. Retrieved from https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, *38*(6), 937-947.

Luong, M. T., & Manning, C. D. (2015). Stanford neural machine translation systems for spoken language domains. In *Proceedings of the 12th International Workshop on Spoken Language Translation: Evaluation Campaign* (pp. 76-79).

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust.* New York: Pantheon.

Massey, G., & Kiraly, D. (2019). The future of translator education: a dialogue. *Cultus: The Journal of Intercultural Mediation and Communication*, *2019*(12), 15-34.

Massey, G., & Wieder, R. (2019). Quality assurance in translation and corporate communications: Exploring an interdisciplinary interface. In *Quality assurance and assessment practices in translation and interpreting* (pp. 57-87). IGI Global.

Mokyr, J. (1990). *The lever of riches: Technological creativity and economic progress.* New York: Oxford University Press.

Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., Bard, N. et al. (2017). Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, *356*(6337), 508-513.

Orlikowski, W. J., & Scott, S. V. (2008). 10 sociomateriality: challenging the separation of technology, work and organization. *Academy of Management annals*, *2*(1), 433-474.

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms: Symbolic actions and change in a regime of knowing. *Organization Science*, *32*(1), 18-41.

Pajarinen, M., & Rouvinen, P. (2014). Computerization threatens one third of Finnish employment. *Etla Brief*, *22*(13.1), 1-6.

Pedersen, D. (2014). Exploring the concept of transcreation–transcreation as "more than translation". *Cultus: The Journal of Intercultural Mediation and Communication*, *7*, 57-71.

Puranam, P. (2018). The Microstructure of Organizations. Oxford: Oxford University Press

Puranam, P. (2021). Human–AI collaborative decision-making as an organization design problem. *Journal of Organization Design*, *10*(2), 75-80.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192-210.

Raj, M., & Seamans, R. (2018). Artificial intelligence, labor, productivity, and the need for firm-level data. In *The economics of artificial intelligence: An agenda* (pp.553-565). Chicago, IL: University of Chicago Press.

Roth, A. E. (2018). Marketplaces, markets, and market design. *American Economic Review, 108*(7), 1609-58.

Schuster, M., Johnson, M., & Thorat, N. (2016, November 22). *Zero-Shot Translation with Google's Multilingual Neural Machine Translation System.* Retrieved from https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html.

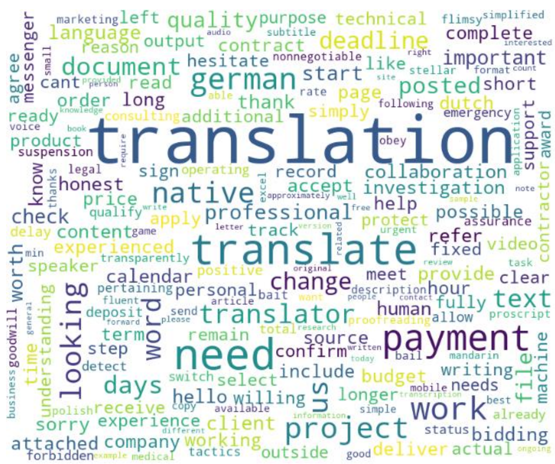
Spender, J. C., & Grant, R. M. (1996). Knowledge and the firm: Overview. *Strategic management journal*, *17*(S2), 5-9.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics*, *24*(2), 235-270.
Statista (2019). *Market size of the global language services industry from 2009 to 2019 with a projection until 2022*. Retrieved from https://www.statista.com/statistics/257656/size-of-the-global-language-services-market/.
Statista (2020). *Number of employees in interpreting and translating services in the United States from 2012 to 2020*. Retrieved from https://www.statista.com/statistics/320340/number-of-employees-in-interpreting-and-translating-services-us.
Sung-won, G. (2017, Feb 14). AI program take on human translators. *The Korea Times*. Retrieved from https://www.koreatimes.co.kr/www/tech/2022/03/133_223942.html?tw.
Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, *42*(9), 1600-1631.
Toews, R. (2021, February 15) *Artificial Intelligence and the end of work.* Retrieved from https://www.forbes.com/sites/robtoews/2021/02/15/artificial-intelligence-and-the-end-of-work/?sh=4051123f56e3.
Tschang, F. T., & Almirall, E. (2021). Artificial intelligence as augmenting automation: Implications for employment. *Academy of Management Perspectives*, *35*(4), 642-659.
Vieira, L. N. (2020). Automation anxiety and translators. *Translation Studies*, *13*(1), 1-21.
Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. *Organization science*, *33*(1), 59-82.
Wang, S., Yang, D. M., Rong, R., Zhan, X., Fujimoto, J., Liu, H. et al. (2019). Artificial intelligence in lung cancer pathology image analysis. *Cancers*, *11*(11), 1673.
Webb, M. (2019). The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.
Wordspath Team. (2023, February 22). ChatGPT Translation Quality: How Is It? Retrieved from https://www.wordspath.com/chatgpt-translation-quality/.
Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W. et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.
Yamada, M. (2019). The impact of Google Neural Machine Translation on Post-editing by student translators. *The Journal of Specialised Translation*, *31*, 87-106.

FIGURES AND TABLES

Figure 1. Word clouds

Figure 1A. Raw data



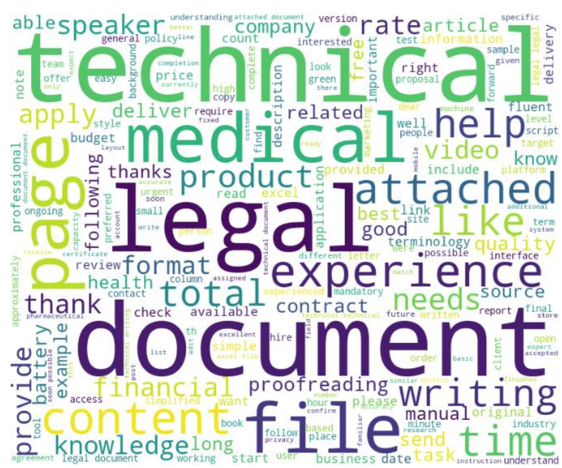
Notes: The plot shows the most common words in the raw data.

Figure 1B. Word cloud for different translation tasks

(1) Regular translations

General content

Technical content

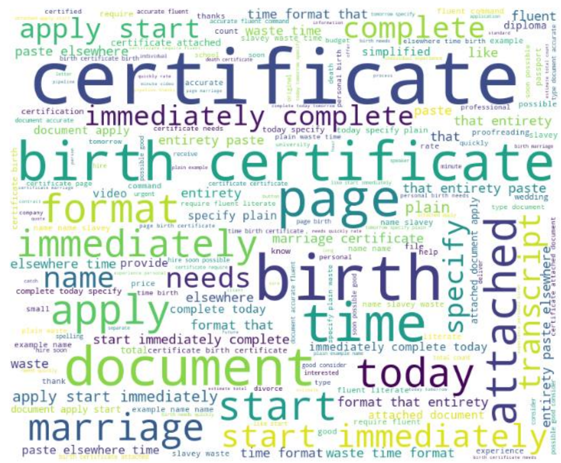


(2) Transcreation

3) Other

Audio/video services

Certificates

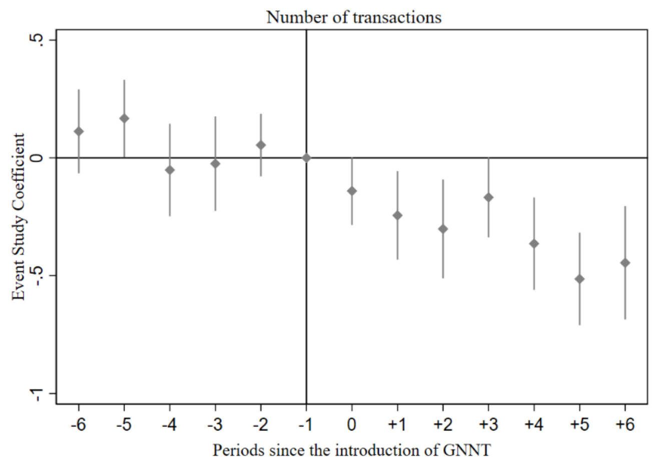


Notes. The plots show the most common words for each of the categories we identified. Note that these word clouds are based on the entire classified sample, with the hand-coded training sample serving as the starting point for the transformer algorithm.

Figure 2. The number of transactions before and after the introduction of GNNT (first cohort of the treatment)

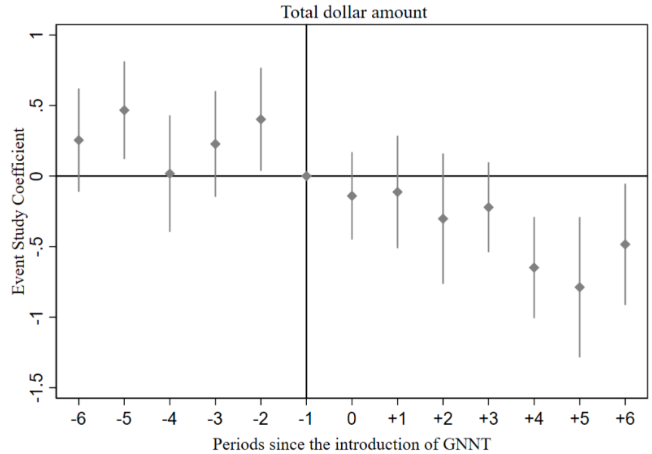
Notes. Transactions are measured in quantity and are normalized to the level in October 2016 (t = −1).

Figure 3. Coefficient estimates on number of transactions before and after the introduction of GNNT.



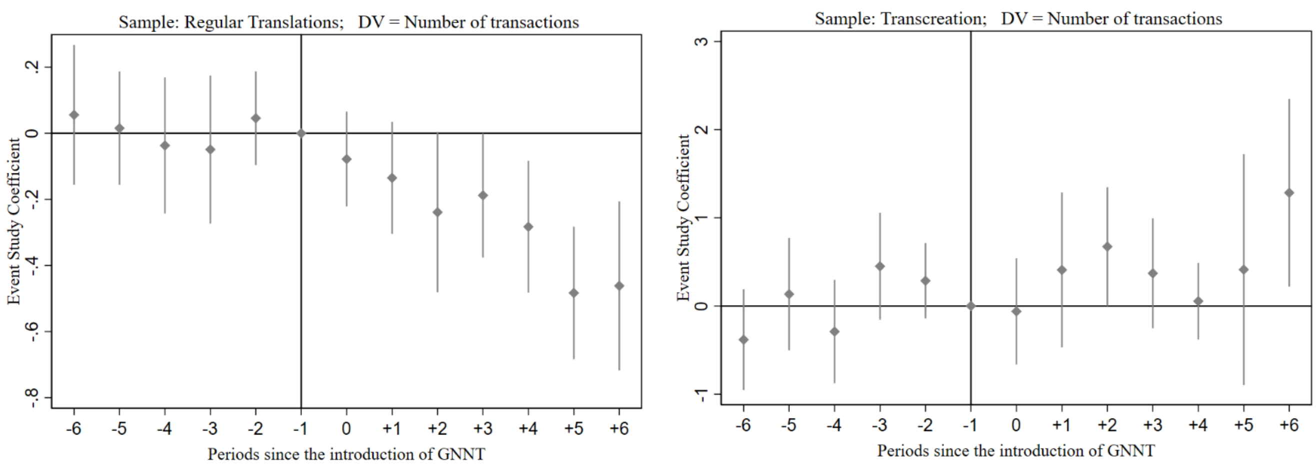
Notes. The vertical axis shows OLS coefficients (and 90 percent confidence intervals). The horizontal axis shows time relative to treatment. The coefficients are obtained from a regression of *Logarithm of Number of Transactions* $= \alpha + \Sigma z\, \beta_1 * 1(z) + \Upsilon i + \delta_t + \varepsilon_{it}$ where Υi is Language-Pair FE, δ_t is Month-year FE, and ε_{it} is the error term. z represents the "lag," or the years relative to a "zero year," which marks the year when *GNNT* first becomes available for a given language pair.

Figure 4. Coefficient estimates on total transaction dollar amount before and after the introduction of GNNT.



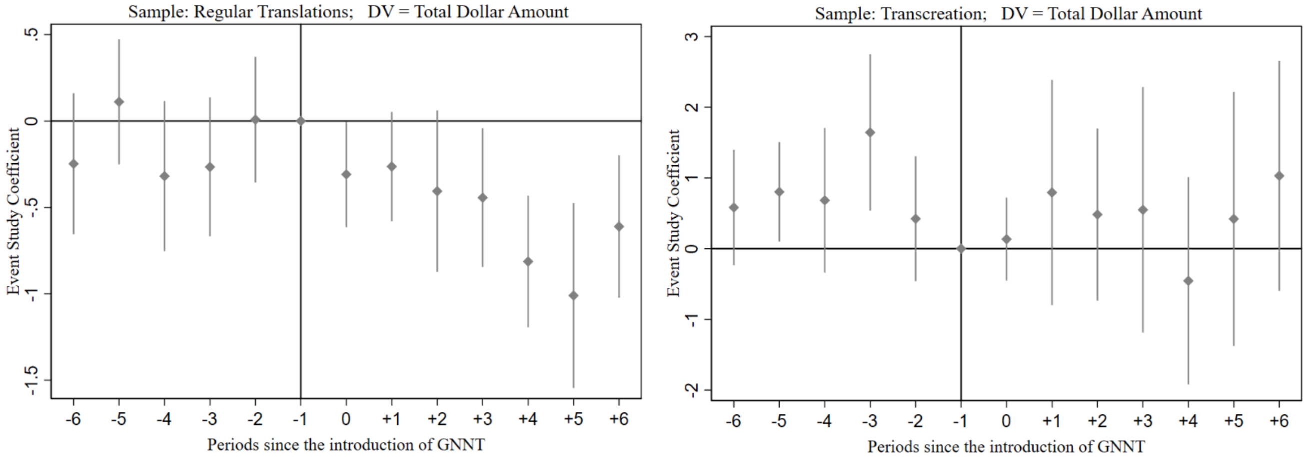
Notes. The vertical axis shows OLS coefficients (and 90 percent confidence intervals). The horizontal axis shows time relative to treatment. The coefficients are obtained from a regression of *Logarithm of Total Dollar Amount* $= \alpha + \Sigma z\ \beta_1 * 1(z) + \Upsilon i + \delta_t + \varepsilon_{it}$ where Υi is Language-Pair FE, δ_t is Month-year FE, and ε_{it} is the error term. z represents the "lag," or the years relative to a "zero year," which marks the year when *GNNT* first becomes available for a given language pair.

Figure 5. Heterogeneous effect of GNNT on different translation tasks (Dependent variable = Logarithm of Number of Transactions)



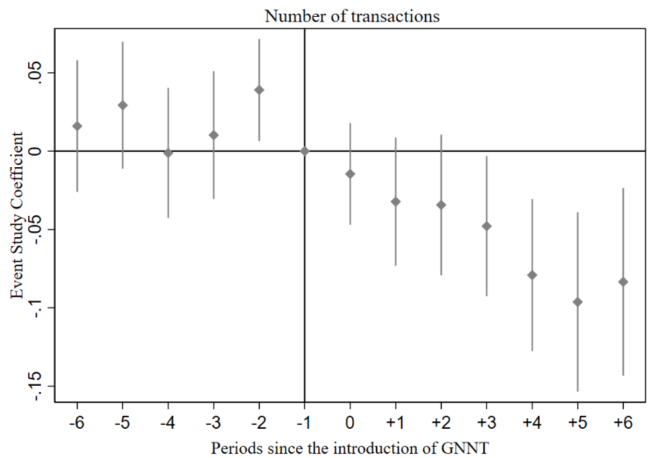
Notes. These graph shows the effect of GNNT on Regular translation and Transcreation tasks separately. The vertical axis shows the OLS coefficient (and 90 percent confidence intervals). The horizontal axis shows time relative to treatment. The coefficients are obtained from a regression of *Logarithm of Number of Transactions* $= \alpha + \Sigma z\ \beta_1 * 1(z) + \Upsilon i + \delta_t + \varepsilon_{it}$ where Υi is Language-Pair FE, δ_t is Month-year FE, and ε_{it} is the error term. z represents the "lag," or the years relative to a "zero year," which marks the year when GNNT first becomes available for a given language pair.

Figure 6. Heterogeneous effect of GNNT on different translation tasks (Dependent variable = Logarithm of Total Dollar Amount)



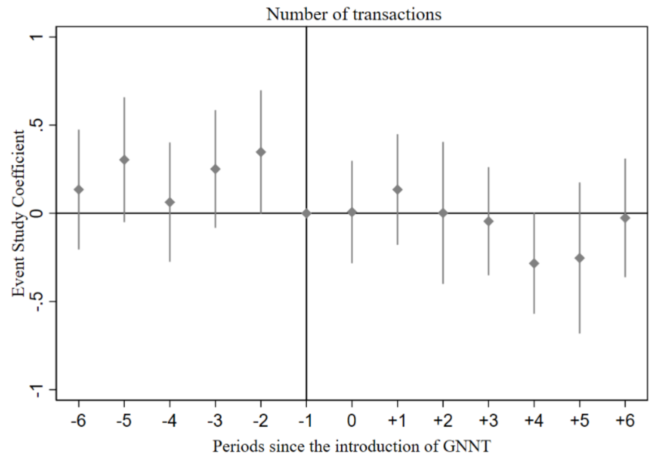
Notes. This graph shows the effect of GNNT on Regular translation and Transcreation tasks separately. The vertical axis shows the OLS coefficient (and 90 percent confidence intervals). The horizontal axis shows time relative to treatment. The coefficients are obtained from a regression of *Logarithm of Total Dollar Amount* $= \alpha + \Sigma z\, \beta_1 * 1(z) + \Upsilon i + \delta_t + \varepsilon_{it}$ where Υi is Language-Pair FE, δ_t is Month-year FE, and ε_{it} is the error term. z represents the "lag," or the years relative to a "zero year," which marks the year when GNNT first becomes available for a given language pair.

Figure 7. Coefficient estimates on transactions at the freelancer level before and after the introduction of GNNT



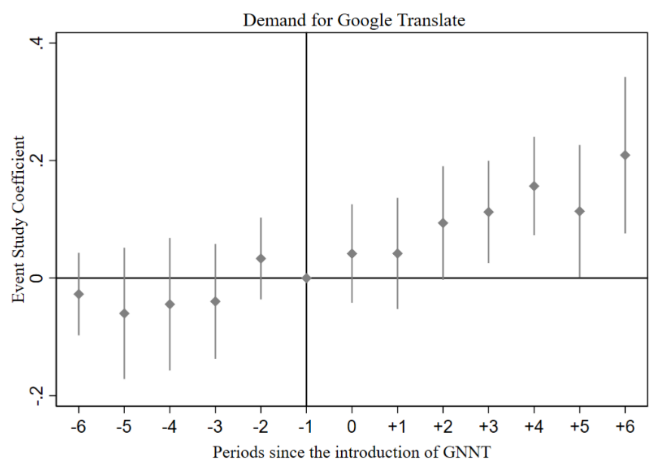
Notes. The vertical axis shows OLS coefficients (and 90 percent confidence intervals). The horizontal axis shows time relative to treatment. The coefficients are obtained from a regression of *Logarithm of Number of Transactions* $= \alpha + \Sigma z\, \beta_1 * 1(z) + \Upsilon i + \delta_t + \varepsilon_{it}$, where Υi is Individual FE, δ_t is Month-year FE, and ε_{it} is the error term. z represents the "lag," or the years relative to a "zero year," which marks the year when GNNT first affects a given freelancer.

Figure 8. Coefficient estimates on the average price of transaction, before and after the introduction of GNNT



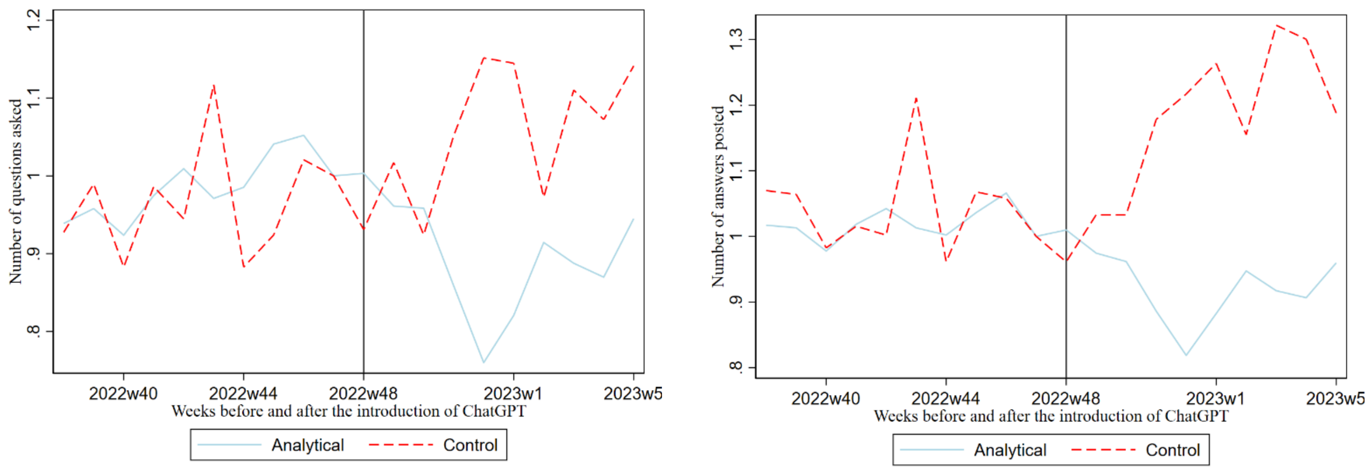
Notes. The vertical axis shows OLS coefficients (and 90 percent confidence intervals). The horizontal axis shows time relative to treatment. The coefficients are obtained from a regression of *Logarithm of Price* $= \alpha + \Sigma z\, \beta_1 * 1(z) + \Upsilon i + \delta_t + \varepsilon_{it}$ where Υi is Language-Pair FE, δ_t is Month-year FE, ε_{it} is the error term. z represents the "lag," or the years relative to a "zero year," which marks the year when GNNT first becomes available for a given language pair.

Figure 9. Coefficient estimates on the demand for Google Translate before and after the introduction of GNNT



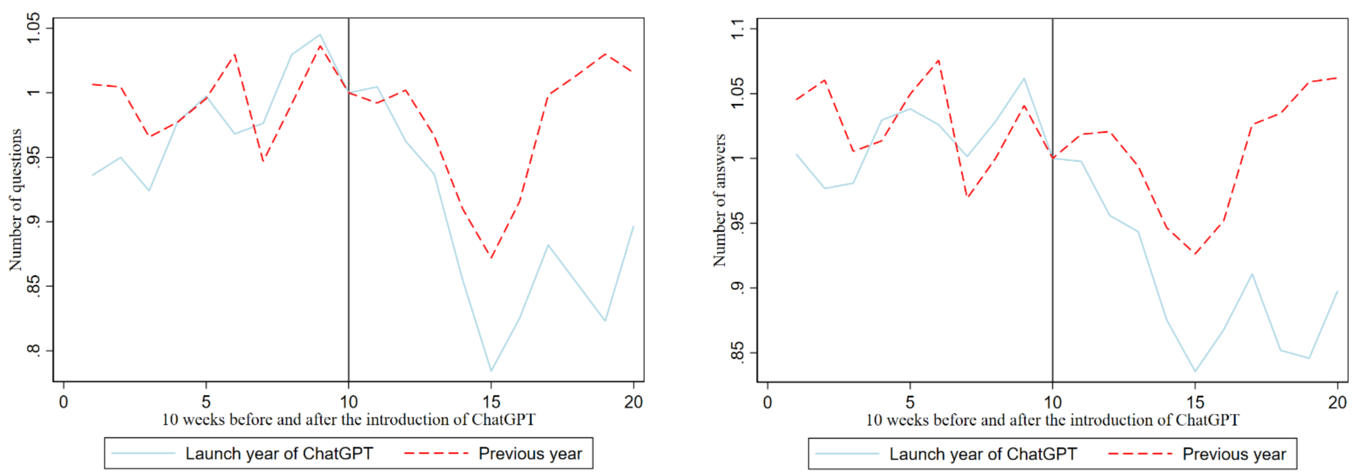
Notes. This graph shows the effect of GNNT on demand for Google Translate. The vertical axis shows OLS coefficients (and 90 percent confidence intervals). The horizontal axis shows time relative to treatment. The coefficients are obtained from a regression of *Logarithm of Query Volumes* $= \alpha + \Sigma z\, \beta_1 * 1(z) + \Upsilon i + \delta_t + \varepsilon_{it}$ where Υi is Individual FE, δ_t is Month-year FE, and ε_{it} is the error term. *z* represents the "lag," or the years relative to a "zero year," which marks the year when GNNT first becomes available for a given language pair.

Figure 10. The number of questions and answers before and after the introduction of ChatGPT



Notes. Transactions are measured in quantity and are normalized to the level in the third week of November 2022 (t = −1).

Figure 11. The number of questions and answers on Stack Overflow compared to last year.



Notes. Transactions are measured in quantity and are normalized to the level in the third week of November in each year (t = −1).

Table 1. Descriptive statistics for language pairs

Variable	Obs	Mean	Std. Dev.	Min	Max
Logarithm of Number of Transactions	782	2.536	1.436	0	5.717
Number of Transactions	782	34.349	53.903	1	304
Logarithm of Total Dollar Amount	782	6.936	1.983	1.253	10.904
Logarithm of Price	782	4.406	.986	1.253	9.09

PostTreatment	782	.147	.354	0	1

Table 2. Baseline Results – Results at the market level

	(I) Logarithm of Number of Transactions	(II) Number of Transactions	(III) Logarithm of Total Dollar Amount
PostTreatment	-0.2283	-0.1342	-0.3853
	(0.0626)	(0.0603)	(0.1316)
	[0.0007]	[0.0259]	[0.0053]
Constant	2.5698	4.3939	6.9924
	(0.0092)	(0.0204)	(0.0194)
	[0.0000]	[0.0000]	[0.0000]
Language Pair FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
N	782	782	782
Adjusted R^2	0.901		0.733
Pseudo-R^2		0.902	

Standard errors are robust to heteroskedasticity, clustered at language pair level and reported in parentheses. p-values are reported in the brackets. Column I and Column III report the result with OLS specification. Column II reports the result from Poisson specification.

Table 3. Heterogeneous effect on different translation tasks – Regular Translation vs. Transcreation

	Regular (I-III)			Transcreation (IV-VI)		
	(I) Logarithm of Number of Transactions	(II) Number of Transactions	(III) Logarithm of Total Dollar Amount	(IV) Logarithm of Number of Transactions	(V) Number of Transactions	(VI) Logarithm of Total Dollar Amount
PostTreatment	-0.1386	-0.1398	-0.2550	0.0656	-0.0231	-0.1874
	(0.0562)	(0.0706)	(0.1330)	(0.1681)	(0.0822)	(0.3630)
	[0.0192]	[0.0477]	[0.0643]	[0.7027]	[0.7791]	[0.6143]
Constant	2.8624	4.2557	7.3288	2.1161	2.8438	6.7063
	(0.0101)	(0.0233)	(0.0240)	(0.0473)	(0.0372)	(0.1022)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Language Pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	561	561	561	238	238	238
Adjusted R^2	0.900		0.738	0.728		0.520
Pseudo-R^2		0.872			0.585	

Standard errors are robust to heteroskedasticity, clustered at language pair level and reported in parentheses. p-values are reported in the brackets. Columns I, III, IV, and VI report the result from OLS specification, and Column II and V report the result from Poisson specification. Please note that the numbers of observations are not exactly equal to main analyses. This is because we restricted our sample to those language pairs that have at least one transaction from the given category.

Table 4. Descriptive statistics for language translation transactions at the individual level

Variable	Obs	Mean	Std. Dev.	Min	Max
Logarithm of Number of Transactions	24371	.403	.613	0	3.401
Number of Transactions	24371	.92	1.941	0	29
PostTreatment	24371	.307	.461	0	1

Table 5. Baseline Results – Results at the individual level

	(I) Logarithm of Number of Transactions	(II) Number of Transactions
PostTreatment	-0.0481	-0.0956
	(0.0284)	(0.0503)
	[0.0906]	[0.0573]
Constant	0.6643	0.6719
	(0.0081)	(0.0145)
	[0.0000]	[0.0000]
Individual FE	Yes	Yes
Month-year FE	Yes	Yes
N	8502	24371
Adjusted R^2	0.384	
Pseudo-R^2		0.392

Standard errors are robust to heteroskedasticity, clustered at individual freelancer level and reported in parentheses. p-values are reported in the brackets. Column I reports the result from OLS specification. Column II reports the result from Poisson specification. The number of observations is slightly lower in the Log-OLS sample because the dependent variable is log-transformed, and any observations with a zero dependent variable are dropped from the sample. The results look identical when we make the log transformation as ln(Number of Transactions + 1)

Table 6. Baseline Results – Effect on Prices

	(I)
	Logarithm of Price
PostTreatment	-0.1493
	(0.0983)
	[0.1360]
Constant	4.4284
	(0.0145)
	[0.0000]
Language Pair FE	Yes
Month-year FE	Yes
N	782
Adjusted R^2	0.188

Standard errors are robust to heteroskedasticity, clustered at language pair level and reported in parentheses. p-values are reported in the brackets.

Table 7. Baseline Results – Demand for Google Translate

	(I) Demand for Google Translate
PostTreatment	0.0856
	(0.0350)
	[0.0185]
Constant	3.9754
	(0.0052)
	[0.0000]
Language Pair FE	Yes
Month-year FE	Yes
Observations	782
Adjusted R^2	0.677

Standard errors are robust to heteroskedasticity, clustered at language pair level and reported in parentheses.
p-values are reported in the brackets.

Table 8. Results at the market level (matched sample)

	(I) Logarithm of Number of Transactions	(II) Number of Transactions	(III) Logarithm of Total Dollar Amount
PostTreatment	-0.2339	-0.2184	-0.3760
	(0.0872)	(0.0497)	(0.1892)
	[0.0105]	[0.0000]	[0.0536]
Constant	2.1890	3.5625	6.5359
	(0.0088)	(0.0111)	(0.0190)
	[0.0000]	[0.0000]	[0.0000]
Language Pair FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
N	714	714	714
Adjusted R^2	0.841		0.625
Pseudo-R^2		0.803	

Standard errors are robust to heteroskedasticity, clustered at language pair level and reported in parentheses. p-values are reported in the brackets. Column I and Column III report the result with OLS specification. Column II reports the result from Poisson specification.

Table 9. Empirical Extension: Evidence with ChatGPT

	(1) Logarithm of Number of Questions	(2) Number of Questions	(3) Logarithm of Number of Answers	(4) Number of Answers
Post	0.0342	0.0834	0.0588	0.1105
	(0.0494)	(0.0321)	(0.0967)	(0.0581)
	[0.5083]	[0.0093]	[0.5596]	[0.0573]
Treatment	6.9669	6.7955	6.5301	6.1734
	(0.2180)	(0.1963)	(0.3055)	(0.3238)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
PostTreatment	-0.1295	-0.1763	-0.1551	-0.2054
	(0.0494)	(0.0321)	(0.0967)	(0.0581)
	[0.0307]	[0.0000]	[0.1472]	[0.0004]
Constant	3.3852	3.5575	3.8520	4.2090
	(0.2180)	(0.1963)	(0.3055)	(0.3238)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
N	180	180	180	180
R^2	0.922		0.850	
Pseudo-R^2		0.998		0.996

Standard errors are robust to heteroskedasticity, clustered at forum level and reported in parentheses. p-values are reported in brackets. Column I and Column III report the result with OLS specification.
Column II-IV report the result from Poisson specification.

Table 10. Empirical Extension: Comparing number of questions and answers posted to previous year.

	(1) Logarithm of Number of Questions	(2) Number of Questions	(3) Logarithm of Number of Answers	(4) Number of Answers
Post	-0.0377	-0.0242	-0.0362	-0.0217
	(0.0082)	(0.0065)	(0.0147)	(0.0099)
	[0.0008]	[0.0002]	[0.0313]	[0.0285]
Treatment	0.0191	0.0716	-0.1768	-0.1309
	(0.4815)	(0.4764)	(0.4982)	(0.4945)
	[0.9691]	[0.8806]	[0.7294]	[0.7912]
PostTreatment	-0.0690	-0.0812	-0.0712	-0.1002
	(0.0129)	(0.0136)	(0.0209)	(0.0221)
	[0.0002]	[0.0000]	[0.0059]	[0.0000]
Constant	7.5334	7.8279	7.5534	7.8753
	(0.3243)	(0.3261)	(0.3391)	(0.3396)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
N	240	240	240	240
Adjusted R^2	0.001		0.019	
Pseudo-R^2		0.003		0.015

Standard errors are robust to heteroskedasticity, clustered at forum level and reported in parentheses. p-values are reported in brackets. Column I and Column III report the result with OLS specification.
Column II-IV report the result from Poisson specification.

APPENDIX

Details of Sample Construction

We were interested in the impact of AI-based translation technology on the market for language pairs that were subject to GNNT, relative to those that were not. Therefore, we first needed to identify the transactions related to a translation, as a freelancer might provide services different from translation and this information is sometimes not explicit. To do so, we first examined the skills required by the clients to complete the job. Specifying the skills requires freelancers to choose from predefined standardized tags, which represent the "controlled vocabulary" of skills maintained by the online marketplace (Horton & Tambe, 2020). In 35% percent of the transactions in our data, clients specify the skills required. For instance, if a transaction involves the translation of content from English to Spanish, the client might choose *"translation-english-spanish."* For the remaining transactions, we first stemmed each word in the job title and description—i.e., reduced them to their root form. For instance, the root forms of the words *"translation," "translated," "translate,"* would be *"translat,"* and for the words *"interpretation," "interpret," "interpreted,"* would be *"interpret."* Then we searched for three stemmed keywords: *"translat,"* "*interpret,"* and *"local"* (the root form of the word "localization," which is sometimes used as a synonym for translation). Using those three keywords, we marked transactions as translation related. Apart from relying on keywords, we also examined different patterns in the data, as clients might refer to translation without using explicit keywords. For instance, some clients just write: *"Articles from English to Italian," "transcribe German to Japanese,"* or *"Subtitles en > fr."* Although 97.2 percent of the transactions were in English, there were some transactions for which descriptions were written in foreign languages. We also determined language pairs by searching these patterns in foreign languages. Some translation projects involve translation into multiple languages and hence require more than one freelancer. Therefore, in those transactions, clients list more than one language pair. To identify the language that freelancers translated, we examined the freelancers' skill sets. In some cases, clients would just write *"Translation"* without specifying the language pairs. In those cases, we looked at the freelancers' skill sets to identify the language pair. If freelancers could translate from and into more than one language pair, and if we could not identify the language pair from the text of the transaction, we dropped those translation transaction from the data (26 transactions). Our final dataset consists of 28,158 transactions between January 2016 and May 2017. We end the sampling period in May 2017 because Google Translate began using neural machine translation for most of the other languages starting in June 2017. Table A1 shows all the languages covered by Google Translate at the end of April—the list remained the same until June 15th. Our criterion for treatment and control selection is having at least one transaction in all of the month-year observations.

Table A1. The list of language pairs affected by GNNT

Language Pair	
Afrikaans <-> English	af <-> en
Albanian <-> English	sq <-> en
Arabic <-> English	ar <-> en
Bulgarian <-> English	bg <-> en
Bengali <-> English	bn <-> en
Chinese (Simplified) <-> English	zh-CN * <-> en
Croatian <-> English	hr <-> en
Czech <-> English	cs <-> en
Danish <-> English	da <-> en
Dutch <-> English	nl <-> en
Finnish <-> English	fi <-> en
French <-> English	fr <-> en
German <-> English	de <-> en
Greek <-> English	el <-> en
Gujarati <-> English	gu <-> en
Hebrew <-> English	iw <-> en
Hindi <-> English	hi <-> en
Hungarian <-> English	hu <-> en
Icelandic <-> English	is <-> en
Indonesian <-> English	id <-> en
Italian <-> English	it <-> en
Japanese <-> English	ja <-> en
Kannada <-> English	kn <-> en
Korean <-> English	ko <-> en
Malayalam <-> English	ml <-> en
Marathi <-> English	mr <-> en
Norwegian <-> English	no <-> en
Polish <-> English	pl <-> en
Portuguese <-> English	pt <-> en
Punjabi <-> English	pa <-> en
Romanian <-> English	ro <-> en
Russian <-> English	ru <-> en
Slovak <-> English	sk <-> en
Spanish <-> English	es <-> en
Swedish <-> English	sv <-> en
Tamil <-> English	ta <-> en
Telugu <-> English	te <-> en
Thai <-> English	th <-> en
Turkish <-> English	tr <-> en
Ukrainian <-> English	uk <-> en
Vietnamese <-> English	vi <-> en

Note: This information was collected initially from the Google AI blog (https://ai.googleblog.com/) and, starting from April 29th, 2017, maintained from https://cloud.google.com/translate/docs/languages

Additional Analyses
Sample statistics for heterogeneous effect regression sample

Table A2. Descriptive statistics for regular translation and transcreation transactions

	N	mean	sd	min	max
Regular Translation					
Logarithm of Number of Transactions	561	2.837	1.312	0	5.451
Number of Transactions	561	36.93	46.592	1	233
Logarithm of Total Dollar Amount	561	7.283	1.775	1.099	10.813
PostTreatment	561	.18	.385	0	1
Transcreation					
Logarithm of Number of Transactions	238	2.135	.96	0	4.043
Number of Transactions	238	12.664	11.243	1	57
Logarithm of Total Dollar Amount	238	6.654	1.541	1.34	9.942
PostTreatment	238	.282	.451	0	1

The vague status of the Chinese-English language pair

According to Mike Schuster, the former head of Google Brain Team, GNNT became available for Chinese-to-English translation in September 2016, but only for translation from English to Chinese in November 2016. In this exercise, we check whether an alternative treatment starting date for the Chinese-English pair affects our results. We report the market-level results on Table A3 and individual-level results on table A4. The results are qualitatively similar to the baseline results.

Table A3. Market-level analysis with alternative sample – Treatment period starts for Chinese-English pair in September 2016

	(I) Logarithm of Number of Transactions	(II) Number of Transactions	(III) Logarithm of Total Dollar Amount
PostTreatment	-0.2153	-0.1058	-0.3701
	(0.0596)	(0.0520)	(0.1269)
	[0.0008]	[0.0419]	[0.0055]
Constant	2.5684	4.3853	6.9911
	(0.0089)	(0.0181)	(0.0190)
	[0.0000]	[0.0000]	[0.0000]
Language Pair FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
N	782	782	782
Adjusted R^2	0.901		0.733
Pseudo-R^2		0.902	

Standard errors are robust to heteroskedasticity, clustered at language pair level and reported in parentheses. p-values are reported in the brackets. Column I and Column III report the result with OLS specification. Column II reports the result from Poisson specification.

Table A4. Individual-level analysis with alternative sample – Treatment period starts for Chinese-English pair in September 2016

	(I) Logarithm of Number of Transactions	(II) Number of Transactions	(III) Logarithm of Number of Transactions	(IV) Number of Transactions
PostTreatment	-0.0485	-0.0485	-0.0574	-0.5233
	(0.0270)	(0.0270)	(0.0852)	(0.2183)
	[0.0730]	[0.0730]	[0.5008]	[0.0165]
			0.0089	0.4326
			(0.0843)	(0.2167)
			[0.9156]	[0.0458]
Constant	0.6673	0.6711	0.6673	0.6710
	(0.0079)	(0.0144)	(0.0079)	(0.0144)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Individual FE	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes
N	8459	24174	8459	24174
Adjusted R^2	0.387		0.387	
Pseudo-R^2		0.390		0.390

Standard errors are robust to heteroskedasticity, clustered at individual freelancer level and reported in parentheses. p-values are reported in the brackets, Column I reports the result from OLS specification. Column II reports the result from Poisson specification. It is important to note that the number of observations is slightly lower, because when we use an alternative treatment starting date for Chinese-English pair some freelancers are dropped from the sample. The number of observations is slightly lower in the Log-OLS sample because the dependent variable is log-transformed, and any observations with a zero dependent variable are dropped from the sample.

Avoiding Staggered Adoption

When a treatment is staggered, it is common practice to conduct a two-way fixed effects difference-in-differences regression to evaluate the treatment effect. However, the estimated average treatment effect may be biased if the treatment effects vary among groups and across time, as they often do in staggered treatment scenarios (Baker et al., 2022). To avoid staggered adoption, we only included observations up to the point at which the second and third cohorts of treatment remained in the control group. We report the results from this exercise on Tables A5 and A6.

Table A5. Baseline Results – Results at the market level

	(I) Logarithm of Number of Transactions	(II) Number of Transactions	(III) Logarithm of Total Dollar Amount
PostTreatment	-0.2758	-0.1252	-0.4260
	(0.0723)	(0.0758)	(0.1469)
	[0.0004]	[0.0987]	[0.0057]
Constant	2.4636	4.3194	6.8867
	(0.0036)	(0.0137)	(0.0073)
	[0.0000]	[0.0000]	[0.0000]
Language Pair FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
N	644	644	644
Adjusted R^2	0.905		0.722
Pseudo-R^2		0.900	

Standard errors are robust to heteroskedasticity, clustered at language pair level and reported in parentheses. p-values are reported in the brackets. Column I and Column III report the result with OLS specification. Column II reports the result from Poisson specification.

Table A6. Heterogeneous effect on different translation tasks – Regular Translation vs. Transcreation

	Regular (I-III)			Transcreation (IV-VI)		
	(I) Logarithm of Number of Transactions	(II) Number of Transactions	(III) Logarithm of Total Dollar Amount	(IV) Logarithm of Number of Transactions	(V) Number of Transactions	(VI) Logarithm of Total Dollar Amount
PostTreatment	-0.2663	-0.1646	-0.5642	0.1910	0.0881	-0.2579
	(0.0745)	(0.0838)	(0.1419)	(0.1824)	(0.0982)	(0.4797)
	[0.0011]	[0.0496]	[0.0004]	[0.3141]	[0.3694]	[0.5998]
Constant	2.7868	4.2010	7.2808	2.0099	2.6538	6.5839
	(0.0052)	(0.0147)	(0.0098)	(0.0261)	(0.0261)	(0.0685)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Language Pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	462	462	462	196	196	196
Adjusted R^2	0.897		0.729	0.730		0.526
Pseudo-R^2		0.871			0.540	

Standard errors are robust to heteroskedasticity, clustered at language pair level and reported in parentheses. p-values are reported in the brackets. Columns I, III, IV, and VI report the result from OLS specification, and Column II and V report the result from Poisson specification. Please note that the numbers of observations are not exactly equal to main analyses. This is because we restricted our sample to those language pairs that have at least one transaction from the given category.