# Don't Just Translate, Summarize Too: Cross-lingual Product Title Generation in E-commerce

**Bryan Zhang, Taichi Nakatani, Daniel Vidal Hussey, Stephan Walter, Liling Tan**

Amazon

{bryzhang,taichina,hussed,sstwa,lilingt}@amazon.com

**Abstract**

Making product titles informative and concise is vital to delighting e-commerce customers. Recent advances have successfully applied monolingual product title summarization to shorten lengthy product titles. This paper explores the cross-lingual product title generation task that summarizes and translates the source language product title to a shortened product title in the target language. Our main contributions are as follows, (i) we investigate the optimal product title length within the scope of e-commerce localization, (ii) we introduce a simple yet effective data filtering technique to train a length-aware machine translation system and compare it to a publicly available LLM, (iii) we propose an automatic approach to validate experimental results using an open-source LLM without human input and show that these evaluation results are consistent with human preferences.

**Keywords:** E-commerce, Summarization, Machine Translation, Natural Language Generation

## 1. Introduction

With e-commerce shopping websites being localized worldwide, products are accessible in different languages through worldwide stores. Moreover, customers are provided with options to browse products in their preferred language other than the primary language of the store. To accomplish this, modern e-commerce stores enable multi-lingual product discovery (Rücklé et al., 2019; Nie, 2010; Saleh and Pecina, 2020; Bi et al., 2020; Jiang et al., 2020; Lowndes and Vasudevan, 2021) as well as localizing product information such as titles using machine translation (MT) systems (Way, 2013; Guha and Heger, 2014; Zhou et al., 2018; Wang et al., 2021).

Localized catalogs (e.g. Amazon, Walmart) contain a large number of products with lengthy titles which are often difficult to read or exceed screen size limits (Zhang et al., 2021; Rozen et al., 2021). This can lead to poor customer experience, especially when titles are used in other contexts such as being read aloud by voice assistants. One reason why localized titles are lengthy is due to their source title: 65% of product titles contain 15 or more words (Rozen et al., 2021) and often intentionally lengthened by online sellers by including redundant keywords and additional product attributes for the purpose of search engine optimization (Xiao and Munro, 2019). Additionally, title length can increase during translation depending on the language pair. Thus, an additional step is required to optimally localize the lengthy title by adhering to the *Grice's maxim of quantity*, i.e. to be informative as required and no more, no less. We hereby refer to this task as "cross-lingual product title generation" (CPTG).

Below is an example of a CPTG task, where a English product title is optimized to a succinct form in Spanish. In the Spanish translation, redundant keywords and extraneous product attributes from the source title are removed.

***`Input:`*** *Rainberg 32cm Frying Pan, Granite Frying Pan Nonstick Coating, Anti-Scratch Pans, Non-Stick Frying Pans, Stone Frying Pan, Induction Compatible, Best Christmas Present, Gift Pack Box. (32cm)*

***`Output:`*** *Rainberg Sartén de granito con revestimiento antiadherente (32 cm)*

In an industry setting, cross-lingual product title generation (CPTG) typically consist of length optimization and localization as two separate steps: 1. Length optimization employs techniques such as monolingual summarization (Sun et al., 2018; Fetahu et al., 2023), text truncation (Wang et al., 2020; Guan et al., 2022) and manual editing. 2. Localization uses machine translation either before or after length optimization.

It is desirable to combine these two steps into one for this CPTG task to reduce operational cost and overhead. To our knowledge the task of jointly accomplishing summarization and machine translation for product titles in e-commerce has not been explored. As neural machine translation (NMT) systems are common in production for localization in e-commerce, it will be convenient to make NMT accomplish both translation and title length optimization in one step. This has the benefit of reducing business complexity and hosting costs.

Therefore in this paper, we first analyse the product title length change during localization. Second, we propose a simple yet effective data filtering technique to train NMT to be aware of product title length optimization. Moreover, multilingual large language models (LLMs) have recently shown promising results on machine translation and summarization tasks and have shown potential to perform both summarization and translation as a single task for certain language pairs. Thus, we also investigate jointly summarizing and translating e-commerce product titles using Large Language Models (LLMs). To validate the effectiveness of using LLMs for the CPTG task, we compare its performance with a smaller NMT system trained for the CPTG task (less than 1 billion parameters). Finally, we propose an novel approach to validate the experimental results using an open source LLM model without human inputs and we show that the evaluation results are consistent with human preferences.

For the rest of the paper, Section 2 discusses the product title length change during localization. Section 3 introduces the architectures we are comparing for the CPTG task. Section 4 proposes an novel way to evaluate experimental results using an LLM with human validation. Section 5 describes the experimental setup. Section 6 presents the results. Section 7 is related work and we conclude in Section 8.

## 2. Product Title Length Change During Localization

We observe that some language pairs have longer product title translations than the source. We sampled hundreds of thousands of source product titles from catalogs, translated them, and calculated the character length ratio of the title translation and source title in Table 1. We find that product title translations tend to be longer for language pairs like English-Spanish and English-German and shorter for pairs such as German-Italian and English-Japanese. The length increase or decrease in the product title translation can be influenced by the target language's grammar, syntax, or even cultural differences. For example, "weekend" in English is "el fin de semana" in Spanish. Thus, in addition to lengthy source product title lengths, their length can also increase during the localization process.

| Language pair with longer title translations | TS Ratio | Language pair with shorter title translations | TS Ratio |
|---|---|---|---|
| English-Spanish | 1.18 | German-Italian | 0.98 |
| German-Polish | 1.08 | German-Swedish | 0.95 |
| English-German | 1.05 | German-Chinese | 0.52 |
| English-Italian | 1.11 | English-Japanese | 0.65 |
| English-Portuguese | 1.12 | English-Arabic | 0.95 |
| English-Polish | 1.16 | French-English | 0.85 |

Table 1: Language pairs with shorter (TS Ratio<1) and longer (TS Ratio>1) product title translations. TS ratio: Length of Target/Length of Source (measured in characters)

## 3. Architectures

### 3.1. Encoder-Decoder Neural Machine Translation with Length Optimization

We propose the following mechanism to select bilingual data and fine-tune an encoder-decoder transformer-based Neural Machine Translation (NMT) model to translate and summarize product information. The data selection approach is to grounded on the length ratio of the target text and source text as follows:

$$R = \frac{L_{tgt}}{L_{src}} \tag{1}$$

$$Select = \begin{cases} 1 & \text{if } R \leq T \\ 0 & \text{otherwise} \end{cases}$$

where $L_{src}$ is the length of the source text, $L_{tgt}$ the length of the target text, $R$ the length ratio, and $T$ as a fixed length ratio threshold.

Intuitively, we select source and target text pairs to fine-tune a machine translation model if the observed length ratio, $R$, is less than a fixed threshold, $T$. This allows the model to gradually learn to generate shorter translations than the source inputs, thereby developing a tendency to optimize length. There are two main advantages of this length ratio-based approach over applying summarization and translation as two steps: (i) it circumvents the need to create bilingual data through pre-summarizing the source input or post-summarizing the translated output, (ii) the length ratio threshold $T$ is adjustable for the given language pair, enabling adaptation to different business requirements.

### 3.2. Prompt-based Cross-lingual Generation using Large Language Models

Prompting has highlighted various emergent capabilities of Large Language Models (LLMs) (Wei et al., 2022a,b; Kojima et al., 2023; Wang et al., 2022). We use `Mixtral-8x7B-Instruct`[1], a publicly available LLM, to translate and summarize the English source product title into the expected target language through prompting using the following *prompt template*. Brackets (<>) are placeholders and substituted with relevant text.

*Prompt Template:*

```
Below is a product title in English:

<source product title text>

Please translate it into <target language> as short
    as possible and in the following JSON format: {
    "output": "<shortened translation>"}.

Do not include other information in the output.
```

[1] https://huggingface.co/mistralai/Mixtral-8x7B-Instruct-v0.1

# 4. Evaluation with LLM and Human Validation

## 4.1. Evaluation with LLM using Test Data without Gold Standard Labels

In an industry setting, proof-of-concept experiments like the CPTG task often do not warrant sufficient resources to be spent on data annotation or purchases. Instead, we propose using the following in-practice evaluation scenario when we first conduct the CPTG task. For this task, we used `Mixtral-8x7B-Instruct`, an open-source LLM chosen for its multilingual abilities, as an evaluator by providing prompts to determine the quality of the CPTG output using the following template.

*Prompt Template:*

```
You are a English to <target language> translation
    expert reviewing product title translations.
Please evaluate if the provided title in <target
    language> is good summary of the English
    product title.

Below is the English title:
<English title>

Below is the title in <target language>:
<title translation>

Here are instructions on how to evaluate:
Respond 2 if the title in <target language>
    translation is shorter than the English title,
    have all of the key product information from
    the English title, and key information in <
    target language> are correct translations.

Respond 1 if the title in <target language>
    translation is shorter than the English title,
    have some of the key product information from
    the English title, and key information in <
    target language> are correct translations.

Respond 0 if the title in <target language>
    translation is shorter than the English title,
    have little of the key product information from
    the English title, and key information in <
    target language> are incorrect translations.

Casing is not important.
Return your response in JSON format  with the 'label
    ' key containing your answer and 'reason' key
    containing a concise justification of your
    answer as {"label": <answer>, "reason": <your
    reasoning>},).

Please make sure the output is in JSON format.
```

And the expected output from the prompt above is as follows:

```
{"label": 1, "reason": <reason>}
```

## 4.2. Validating a Sample of the Test Data with Human Translators

Despite the less than standard approach to test set evaluation in industrial systems as presented in Section 4, we propose to monitor the systems by periodically sampling production traffic for newer experimental systems (Cabrera et al., 2023). This deviates from typical offline evaluation methods where a full gold standard corpus with human annotations is available from the beginning of the experiment. In practice, it is often the regular sub-sampling from production traffic that eventually builds an incremental gold standard corpus over time.[2,3]

# 5. Experimental Setup

**Language pairs**: To compare the architectures described in Section 3, we evaluate them across two language pairs for the CPTG task, English to German (EN-DE) and English to Spanish (EN-ES).

**Encoder-Decoder NMT and LLM-based models:** We refer to the length-optimized fine-tuned Encoder-Decoder NMT model as “Summary”, and the LLM translator as “Mixtral”. A base neural machine translation model (hereby referred to as “Baseline MT”) was used to fine-tune Summarly. Baseline MT consists of a 20 encoder and 2 decoder layer transformer trained using the Sockeye MT toolkit (Hieber et al., 2022) using a large quantity of bilingual generic web data and catalogue product data (product titles, bullet points and description). To fine-tune Summarly, we sampled ~800k EN-ES bilingual product titles with a length ratio $T$ set at 0.7 and ~270k EN-DE segments with $T$ set at 0.8. Development sets of 2000 segments with respective length ratios for the language pairs were created separately for model validation during the fine-tuning process.

**Test set for LLM evaluation and human validation**: We sampled 2000 English source product titles with more than 200 characters from the US store. We chose this length to adhere to common title length guidelines specified on e-commerce sites [4,5,6]. Furthermore, we also sub-sampled 200 segments from the test set obtained two versions of translated titles generated by Summarly and Mixtral. We then provided human raters with these outputs and an evaluation guideline to choose the preferred translation for each language pair.

[2] https://www.splunk.com/en_us/blog/it/building-an-ai-assistant-for-splunk.html

[3] https://www.oreilly.com/library/view/responsible-machine-learning/9781492090878/

[4] https://sellercentral.amazon.com/help/hub/reference/external/GYTR6SYGFA5E3EQC

[5] https://www.ebay.com/sellercenter/listings/listing-best-practices

[6] https://www.etsy.com/seller-handbook/article/366470356778

## 6. Results and Analysis

### 6.1. Length Analysis of Generated Product Titles

Tables 2 and 3 analyze the length of the localized titles from the two models. Both Summarly and Mixtral models are able to translate product titles with shorter lengths. However the Summarly model's optimization is more consistent than Mixtral, with 99% and 92% of test cases being under 200 characters for the two language pairs.

| | Baseline MT | Summarly | Mixtral |
|---|---|---|---|
| Avg. length (in char) | 194 | 155 | **136.6** |
| Shorter (<200 char) % | 52.3% | **92.8%** | 92.2% |
| Shorter (<Baseline MT) % | n/a | **94.1%** | 91.7% |

Table 2: Title generation length analysis for EN-DE

| | Baseline MT | Summarly | Mixtral |
|---|---|---|---|
| Avg. length (in character) | 226 | **127** | 148.8 |
| Shorter (<200 char) % | 2.6% | **99.4%** | 83.1% |
| Shorter (<Baseline MT) % | n/a | **99.9%** | 92.7% |

Table 3: Title generation length analysis for EN-ES

### 6.2. Evaluating Title Quality with an LLM

Table 4 presents the LLM evaluation result. The table represents the number of title translations with respective quality scores assessed by the Mixtral LLM evaluator using the prompt template described in Section 4. Title translations scored with 1 or 2 are considered high-quality title translations, while those scored with 0 are considered low-quality. Across the two language pairs, both Summarly and Mixtral return high-quality summarized titles for over 90% of the test sets. Results show that Summarly generated high-quality title translations for 98% and 93% of the tests set for EN-DE and EN-ES respectively, while Mixtral generated high-quality translations for 93% and 91% of the test set for the two language pairs. Note that the LLM evaluator generated either empty outputs or an invalid JSON output for a subset of evaluation results, ranging between 1.8% and 6.3% across the two test sets.

Examples in the Table 5 show that Summarly not only translated product titles with a more optimized length, but also learned to the summarize. In Example 1, Summarly translates the product title from English to German that are shorter in length while also preserving the key product information. Although the generated product title from Mixtral also preserves key information, its translations are less accurate than Summarly. Example 2 shows Summarly's ability to summarize explicitly for length optimization, where *"...for Kids 3 4 5 6 7 8 Years Old"* in the source title is summarized to *"para niños de 3 a 8 años"* in Spanish.

| | EN-DE | | EN-ES | |
|---|---|---|---|---|
| | **Summarly** | **Mixtral** | **Summarly** | **Mixtral** |
| # Score 0 | 0.05% | 3.35% | 0.65% | 3.55% |
| # Score 1 | 5.15% | 9.65% | 40.45% | 19.75% |
| # Score 2 | 92.95% | 83.40% | 52.60% | 71.90% |
| **High quality (# Score 1 and 2)** | **98.10%** | 93.05% | **93.05%** | 91.65% |
| No results | 1.85% | 3.60% | 6.33% | 4.80% |

Table 4: Evaluation of Title Quality by Mixtral LLM

### 6.3. Human Validation

We sampled 200 translations from Summarly and Mixtral from the test set and asked a human translator to choose the preferred translation for each language pair. Translators were given similar instructions to those used in the prompt template for the Mixtral LLM to evaluate these translations, focusing on translation quality and coverage of key product information coverage. Table 6 presents the results of the human preference of the translations from Summarly and Mixtral. Human validation results were similar to those by the LLM in Table 4, where Summarly's title translations were preferred over Mixtral. This also indicates the validity of the evaluation method described in Section 3.2 to evaluate translation and summarization quality using a LLM in lieu of human evaluations, similar to modern neural machine translation evaluation metrics like GEMBA (Kocmi and Federmann, 2023b,a).

| | Summarly | Mixtral | Tie |
|---|---|---|---|
| EN-DE | **120 (60%)** | 23 (12%) | 57 (28%) |
| EN-ES | **120 (60%)** | 19 (10%) | 60 (30%) |

Table 6: Human validation of test samples

### 6.4. Interpretation of the LLM Evaluation and Human Validation Results

Our analysis of LLM evaluation and human validation (Tables 4 and 6) provide insights into whether title quality evaluations between competing models (Summarly and Mixtral) are consistent between an LLM and human judgment. However, this approach has limitations. While using LLMs to automate quality evaluations of localized product titles by an LLM can provide insights to stakeholders, they may not accurately reflect its assessment across the entire production traffic. Nevertheless, our findings provide a signal to further invest in the industrial capabilities of the CPTG task with the following recommendations: (1) Deploy a baseline model

| | |
|---|---|
| **Example 1 (English to German)** | |
| Source Title | Baking Parchment Paper, Non-Stick Rounds Greaseproof Baking Sheets Cake Pans Circle Cake Tin Liners Burger Discs Extra Thick - 100 Pcs/Pack (7.6cm(3"))) |
| Baseline MT | Backpapier, antihaftbeschichtet, rund, fettdicht, Backbleche, Kuchenformen, runde Kuchenformen, Burgerscheiben, extra dick, 100 Stück/Packung (7,6 cm) |
| Summarly | Backpapier, antihaftbeschichtet, rund, fettdicht, für Kuchenformen, extra dick, 100 Stück/Packung (7,6 cm) |
| Mixtral | Backpapier: 100 Stück/Pack (7,6 cm), nicht-anhaftend, rund, extra dick) |
| **Example 2 (English to Spanish)** | |
| Source Title | Dinosaur Gift Toys - Dinosaur Arts and Crafts Painting kit Including 12 Cute Dinosaur Figures, DIY Creative Toy Gift for Kids 3 4 5 6 7 8 Years Old |
| Baseline MT | Juguetes de regalo de dinosaurio – Kit de pintura de arte y manualidades de dinosaurios que incluye 12 lindas figuras de dinosaurio, juguete creativo de bricolaje para niños de 3, 4, 5, 6, 7, 8 años |
| Summarly | Juguetes de pintura de dinosaurios, incluye 12 figuras de dinosaurios, juguete creativo para niños de 3 a 8 años |
| Mixtral | Juguetes regalo de dinosaurio para niños y niñas a partir de 3 años - Kit de pintura y artesanía de 12 figuras de dinosaurios + juguete creativo DIY |

Table 5: Cross-lingual product title generation examples

to a production setting to test actual customer experience. (2) Develop a competing newer architecture model that can be improved before it can replace the baseline. (3) Deploy an LLM to to monitor production traffic and automatically evaluate localized title quality at scale. (4) Simultaneously, sub-sample localized titles and assessments by the LLM to perform human validation to ensure the evaluation model is aligned with human judgments.

## 7. Related Work

To address the issue of overly lengthy titles, Sun et al. (2018) introduced the task of Product Title Summarization (PTS) to extract a natural representation of the product while retaining key product details. Recent approaches also use instruction-tuned LLMs to summarize product titles (Fetahu et al., 2023). However, these summarization systems generally focus on the summarization task, which are then cascaded with machine translation systems to carry out the localization task in industry settings. Although length optimization has been widely studied in machine translation for video subtitle and speech-related domains (Yang et al., 2020; Lakew et al., 2021, 2022), to the best of our knowledge we are not aware of prior work on length-optimized machine translation specifically for e-commerce product catalogs.

## 8. Conclusion

We discussed the issue of lengthy product title translations which can occur during the localization process in the e-commerce domain, and proposed the Cross-lingual Product Title Generation (CPTG) task to address such cases. We introduced an automatic, prompt-based evaluation method to evaluate CPTG models without human inputs, using an open-source Mixtral model to compare a length-aware, encoder-decoder transformer-based machine translation model (Summarly) against an LLM (Mixtral). Finally, we validated that the automatic prompt-based evaluation method results align with human assessments, showing that localized titles generated by Summarly outputs were overall preferred over those generated by Mixtral.

## 9. Bibliographical References


Tianchi Bi, Liang Yao, Baosong Yang, Haibo Zhang, Weihua Luo, and Boxing Chen. 2020. Constraint translation candidates: A bridge between neural query translation and cross-lingual information retrieval.

Ángel Alexander Cabrera, Marco Tulio Ribeiro, Bongshin Lee, Robert Deline, Adam Perer, and Steven M. Drucker. 2023. What did my AI learn? how data scientists make sense of model behavior. *ACM Trans. Comput.-Hum. Interact.*, 30(1).

Besnik Fetahu, Zhiyu Chen, Oleg Rokhlenko, and Shervin Malmasi. 2023. InstructPTS: Instruction-tuning LLMs for product title summarization. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 663–674, Singapore. Association for Computational Linguistics.

Xinyi Guan, Shun Long, Weiheng Zhu, Silei Cao, and Fangting Liao. 2022. Mask-based text scoring for product title summarization. In *2022 8th International Conference on Systems and Informatics (ICSAI)*, pages 1–6.

Jyoti Guha and Carmen Heger. 2014. Machine translation for global e-commerce on ebay. In *Proceedings of the AMTA*, volume 2, pages 31–37.

Felix Hieber, Michael Denkowski, Tobias Domhan, Barbara Darques Barros, Celina Dong Ye, Xing Niu, Cuong Hoang, Ke Tran, Benjamin Hsu, Maria Nadejde, Surafel Lakew, Prashant Mathur, Anna Currey, and Marcello Federico. 2022. Sockeye 3: Fast neural machine translation with PyTorch.

Zhuolin Jiang, Amro El-Jaroudi, William Hartmann, Damianos Karakos, and Lingjun Zhao. 2020. Cross-lingual information retrieval with BERT. In *Proceedings of the workshop on Cross-Language Search and Summarization of Text and Speech (CLSSTS2020)*, pages 26–31, Marseille, France. European Language Resources Association.

Tom Kocmi and Christian Federmann. 2023a. GEMBA-MQM: Detecting translation quality error spans with GPT-4. In *Proceedings of the Eighth Conference on Machine Translation*, Singapore. Association for Computational Linguistics.

Tom Kocmi and Christian Federmann. 2023b. Large language models are state-of-the-art evaluators of translation quality. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 193–203, Tampere, Finland. European Association for Machine Translation.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2023. Large language models are zero-shot reasoners.

Surafel M. Lakew, Marcello Federico, Yueru Wang, Cong Hoang, Yash Virkar, Roberto Barra-Chicote, and Róbert Enyedi. 2021. Machine translation verbosity control for automatic dubbing. In *Proceedings of IEEE-ICASSP*.

Surafel M. Lakew, Yogesh Virkar, Prashant Mathur, and Marcello Federico. 2022. Isometric MT: Neural machine translation for automatic dubbing. In *Proceedings of IEEE-ICASSP*.

Mike Lowndes and Aditya Vasudevan. 2021. Market guide for digital commerce search.

Jian-Yun Nie. 2010. Cross-language information retrieval. *Synthesis Lectures on Human Language Technologies*, 3(1):1–125.

Ohad Rozen, David Carmel, Avihai Mejer, Vitaly Mirkis, and Yftah Ziser. 2021. Answering product-questions by utilizing questions from other contextually similar products. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 242–253, Online. Association for Computational Linguistics.

Andreas Rücklé, Krishnkant Swarnkar, and Iryna Gurevych. 2019. Improved cross-lingual question retrieval for community question answering. In *The World Wide Web Conference*, WWW '19, page 3179–3186, New York, NY, USA. Association for Computing Machinery.

Shadi Saleh and Pavel Pecina. 2020. Document translation vs. query translation for cross-lingual information retrieval in the medical domain. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6849–6860, Online. Association for Computational Linguistics.

Fei Sun, Peng Jiang, Hanxiao Sun, Changhua Pei, Wenwu Ou, and Xiaobo Wang. 2018. Multi-source pointer network for product title summarization. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 7–16, New York, NY, USA. Association for Computing Machinery.

Haifeng Wang, Hua Wu, Zhongjun He, Liang Huang, and Kenneth Ward Church. 2021. Progress in machine translation. *Engineering*.

Manyi Wang, Tao Zhang, Qijin Chen, and Chengfu Huo. 2020. Selling products by machine: a user-sensitive adversarial training method for short title generation in mobile e-commerce.

Thomas Wang, Adam Roberts, Daniel Hesslow, Teven Le Scao, Hyung Won Chung, Iz Beltagy, Julien Launay, and Colin Raffel. 2022. What language model architecture and pretraining objective work best for zero-shot generalization?

Andy Way. 2013. Traditional and emerging use-cases for machine translation. *Proceedings of Translating and the Computer*, 35:12.

Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. 2022a. Fine-tuned language models are zero-shot learners.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. 2022b. Emergent abilities of large language models.

Joan Xiao and Robert Munro. 2019. Text summarization of product titles. In *eCOM@SIGIR*.

Zijian Yang, Yingbo Gao, Weiyue Wang, and Hermann Ney. 2020. Predicting and using target length in neural machine translation. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 389–395, Suzhou, China. Association for Computational Linguistics.

Xueying Zhang, Yunjiang Jiang, Yue Shang, Zhaomeng Cheng, Chi Zhang, Xiaochuan Fan, Yun Xiao, and Bo Long. 2021. DSGPT: Domain-specific generative pre-training of transformers for text generation in e-commerce title and review summarization. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 2146–2150, New York, NY, USA. Association for Computing Machinery.

Mingyang Zhou, Runxiang Cheng, Yong Jae Lee, and Zhou Yu. 2018. A visual attention grounding neural model for multimodal machine translation. *CoRR*, abs/1808.08266.