

# HGP-NLP at BioLaySumm: Leveraging LoRA for Lay Summarization of Biomedical Research Articles using Seq2Seq Transformers

Hemang Malik<sup>\*†</sup> Gaurav Pradeep<sup>‡†</sup> Pratinav Seth<sup>\*</sup>

Manipal Institute of Technology

Manipal Academy of Higher Education, Manipal, India

{hemangmalik2486, gauravpradeep2004, seth.pratinav}@gmail.com

## Abstract

Lay summarization aims to generate summaries of technical articles for non-experts, enabling easy comprehension for a general audience. The technical language used in research often hinders effective communication of scientific knowledge, making it difficult for non-experts to understand. Automatic lay summarization can enhance access to scientific literature, promoting interdisciplinary knowledge sharing and public understanding. This has become especially important for biomedical articles, given the current global need for clear medical information. Large Language Models (LLMs) (Yao et al., 2024), with their remarkable language understanding capabilities, are ideal for abstractive summarization, helping to make complex information accessible to the public. This paper details our submissions to the BioLaySumm 2024 Shared Task: Lay Summarization of Biomedical Research Articles (Goldsack et al., 2024). We fine-tune and evaluate sequence-to-sequence models like T5 across various training dataset settings and optimization methods such as LoRA (Hu et al., 2021b) for lay summarization. Our submission achieved the 53rd position overall.

## 1 Introduction

Scientific research aims to advance knowledge, but a major challenge is the lack of domain-specific knowledge among the public. Technical reports and research articles are often incomprehensible to non-experts, hindering knowledge dissemination. Lay summarization addresses this by generating factual, readable summaries of technical texts for non-experts (Chaturvedi et al., 2020). In the biomedical field, lay summarization is crucial due to its highly technical content involving complex medical terms and detailed research findings.

Access to clear medical information is essential for informed health decisions. Misunderstandings or a lack of access can lead to poor decisions, increased anxiety, and a general distrust of medical advice (Guo et al., 2021). Recent advancements in large language models (LLMs) (Yao et al., 2024) and autoregressive LLMs (Chen et al., 2023) like ChatGPT, Gemini, Mistral, and Llama have shown significant potential. With their large-scale pre-training, these models can generate high-quality, contextually relevant summaries that are informative and accessible. However, they are computationally expensive and often require fine-tuning for specific tasks, needing substantial computational resources and energy. Sequence-to-sequence models (Chiu et al., 2018) provide a promising alternative, addressing these computational challenges. With an encoder-decoder structure, these models efficiently handle input and output sequences, making them well-suited for summarization (Kouris et al., 2021). They are more computationally efficient than LLMs, requiring fewer resources while delivering high performance. This paper describes our approach to the BioLaySumm 2024 Shared Task: Lay Summarization of Biomedical Research Articles (Goldsack et al., 2024). We fine-tune and evaluate sequence-to-sequence models like T5 for the task of lay summarization (Challagundla and Peddavenkatagari, 2024) across various training dataset settings and optimization methods such as LoRA (Hu et al., 2021b). Our final submission to this task is a LoRA-based Flan-T5-Base model (Rusnachenko and Liang, 2024), which secured us the 53rd position on the leaderboard.

## 2 Background and Previous Works

### 2.1 Problem and Data Description

The BioLaySumm shared task at the BioNLP Workshop of ACL 2024 (Goldsack et al., 2024) focuses on the abstractive summarization of biomedical ar-

\* Dept. of Data Science & Computer Applications

† Equal Contribution

‡ Dept. of Computer Science & Engineering

Attributes	Values
Article	In the USA , more deaths happen...
Headings	Abstract, Introduction...
Keywords	epidemiology and global health...
Id	elife-35500-v1...

Table 1: Text Samples from eLife Training dataset with their corresponding attributes

ticles. The goal is to generate summaries that are accessible to non-expert audiences by including more background information and using less technical terminology, effectively creating a “lay summary.” Task Definition: Given an article’s abstract and main text, participants are required to train a model (or models) to generate a lay summary. Two separate datasets, derived from the biomedical journals PLOS and eLife, are provided for model training and will be used for evaluation. For the final evaluation, submissions will be ranked based on their average performance across both datasets.

BioLaySumm offers two datasets from well-respected scientific journals: the Public Library of Science (PLOS) and eLife. Each dataset provides pairs of original research articles along with their corresponding summaries, written by scientists specifically for a public audience, as indicated in Table 1. The PLOS dataset is the larger of the two, containing 24,773 instances for training, 142 for testing, and 1,376 for validation. The eLife dataset contains 4,346 instances for training, 142 for testing, and 241 for validation, as shown in the Table 2. Each dataset includes train, validation, and test sets in the form of JSONL files. The eLife summaries have an average length of around 300-350 words, whereas the PLOS summaries are shorter, averaging around 160-200 words (Goldsack et al., 2022).

## 2.2 Automatic Text Summarization

Summarization involves condensing large amounts of text into key points. Extractive summarization, like underlining key passages in a textbook, extracts important sentences directly from the original text. This method is simpler and ensures factual accuracy, but the summaries may lack coherence (Neto et al., 2002). On the other hand, Abstractive summarization aims to understand and rephrase the main ideas in a new, concise way, offering more readable summaries but potentially introducing errors if the model misinterprets the original text (Gupta and Gupta, 2019). It goes beyond copy-

Dataset	Training	Testing	Validation
PLOS	24,773	142	1,376
eLife	4,346	142	341

Table 2: Frequency of Task labels in dataset

ing sentences, striving to grasp the core message of a text. Traditional summarization approaches typically fall into two categories: extractive methods that select and combine important sentences from the source text, and abstractive methods that generate entirely new text that captures the main ideas. This field has seen significant progress with transformer-based methods excelling at understanding complex relationships in text. Recently, large language models (LLMs) (Alberts et al., 2023) utilizing autoregressive techniques have emerged as a powerful approach, offering the ability to not only summarize but also generate creative text formats (Poornash et al., 2023).

## 2.3 Biomedical Lay Summarization

### 2.3.1 Biomedical Summary Corpora

In the field of deep learning, large collections of text, known as corpora (Stubbs, 2004), are crucial for training language models. These datasets help machines learn the patterns, meaning, and structure of language, thereby improving their ability to understand and generate text for tasks such as sentiment analysis or machine translation (Gilquin and Gries, 2009). Instead of creating a new corpus from scratch, focus on utilizing existing collections of scientific articles alongside their corresponding, simplified summaries written for the general public. The BioLaySumm datasets (PLOS and eLife) are built by extracting these paired texts from the corpora. These datasets then become the foundation for training and evaluating models that can automatically generate clear summaries of scientific research articles for a non-scientific audience (Goldsack et al., 2022).

### 2.3.2 Sequence-to-sequence models

Significant research has utilized sequence-to-sequence (seq2seq) models (Sriram et al., 2017) for generating lay summaries of biomedical research articles. Leveraging pre-trained models like T5 and BART (Colak and Karadeniz, 2023), researchers have explored methods to effectively capture the complexities of biomedical language and translate it into clear summaries for the public. Another area

of focus involves adapting Transformers (Dandan et al., 2023) by incorporating biomedical domain knowledge, such as using pre-trained models on scientific text or knowledge graphs, to enhance the model's understanding of scientific concepts. Additionally, research (Fan et al., 2018) investigates controllable summarization with seq2seq models, allowing researchers to tailor summaries to specific needs like factual accuracy or readability (Sriram et al., 2017).

### 2.3.3 Generative Pre-trained Transformers

Generative Pre-trained Transformers (GPT) (Zhu and Luo, 2022) are advanced auto-regressive transformer models trained on vast amounts of data, enabling them to generate human-like text and perform various natural language understanding tasks with high accuracy and efficiency (Luo et al., 2022). Their strong performance is attributed to the quality of the training dataset and their autoregressive nature, where the input serves as the "prefix" of the output. However, to maximize their performance, fine-tuning on specific downstream tasks is necessary, which can be computationally expensive.

### 2.3.4 Infusing External Knowledge

Integrating external knowledge into models (Paulheim, 2017) involves incorporating domain-specific data or structured information, which boosts performance and contextual comprehension for more accurate predictions and insights. This approach enriches deep learning models (Menghani, 2023) with additional context and expertise, enhancing their robustness and interpretability (Koncel-Kedziorski et al., 2022). One widely used technique involves integrating external knowledge graphs (Paulheim, 2017) for downstream tasks like summarization. This leverages semantic relationships to better understand and contextualize concepts within summaries (Goldsack et al., 2023).

## 2.4 Current State of NLP

State-of-the-art foundation large language models (LLMs) like ChatGPT (Firat, 2023), Gemini (Khurdula et al., 2024), and LLaMA (Touvron et al., 2023) are revolutionizing how we interact with text. These LLMs are auto-regressive in nature and can generate various creative text formats and translate languages with impressive fluency. However, they still require fine-tuning to achieve state-of-the-art performance in downstream tasks. This can be done by optimizing on a downstream dataset

and task or by using methods involving prompting. Fine-tuning these models is computationally expensive due to their large number of parameters. Techniques such as Low-Rank Adaptation of Large Language Models (LoRA) (Hu et al., 2021b) and Parameter-Efficient Fine-Tuning (PEFT) (Sabry and Belz, 2023) address this challenge. These methods enable fine-tuning by freezing the pre-trained model weights and injecting trainable rank decomposition matrices into each layer of the Transformer architecture, greatly reducing the number of trainable parameters for downstream tasks.

## 3 System Overview

Our approach to the task involved fine-tuning sequence-to-sequence transformer-based models (Ramachandran et al., 2016), such as T5, across various training dataset settings and optimization methods like LoRA on the PLoS and eLife datasets. The models were evaluated based on the readability (Paasche-Orlow et al., 2003), relevance (Montague and Aslam, 2001), and factuality (Pagnoni et al., 2021) of the summaries generated. The model demonstrating the best performance across these three evaluation criteria was chosen as our final submission for the task.

### 3.1 Data Pre-Processing

Each sample in the dataset has attributes as illustrated in Table 1. These were utilized to design a prompt structured as follows:

```
Provide a lay summary of the following
article, which includes keywords <key-
words>: <article>.
```

For processing text, the T5 tokenizer was used with a maximum input of 1024 tokens. LoRA-based models, on the other hand, could handle up to 2048 tokens as input (Seye et al., 2018). In all cases, the output was capped at 512 token.

### 3.2 Fine-Tuning Transformers

T5 is a text-to-text Transformer architecture that treats all tasks, like translation and classification, as generating target text from input text. It differs from BERT by adding a causal decoder and using various pre-training tasks instead of the cloze task. We trained the sequence-to-sequence transformer T5 on the given datasets using different approaches. This included training solely on either the eLife or PLoS dataset, followed by inference on the evaluation and test sets. Alternatively, we also trained

Model Name	Data Type	Relevance				Readability			Factuality		
		R-1	R-2	R-L	BTs	FL	DS	CI	LS	As	Sc
T5-small	1	<b>0.297</b>	0.086	<b>0.269</b>	0.837	11.201	9.913	12.786	44.223	0.794	0.741
	2	0.296	0.087	0.268	<b>0.838</b>	<b>11.351</b>	<b>10.158</b>	<b>13.135</b>	<b>48.178</b>	<b>0.840</b>	<b>0.766</b>
LORA Flan T5-base	1	0.296	<b>0.087</b>	0.268	<b>0.838</b>	<b>11.351</b>	<b>10.158</b>	<b>13.135</b>	<b>48.178</b>	<b>0.840</b>	<b>0.766</b>

Table 3: Results of Experimented Models on the Test Set. ‘Data type’ column indicates the number of models used: -1 (unified) or 2 (one for each dataset). Here, R=ROUGE F1, BTs=BERTScore, FL=Flesch-Kincaid Grade Level, DS=Dale-Chall Readability Score, CI=Coleman-Liau Index, LS=LENS metric, As=AlignScore, and Sc=SummaC.

the model on a combined dataset consisting of both train sets before performing inference. Flan-T5 is an extension of T5 (Rusnachenko and Liang, 2024) that has been further fine-tuned on a diverse set of instructions, enhancing its performance on various downstream tasks, including summarization. To fine-tune this model, we employed the Parameter Efficient Fine Tuning Method, specifically the Low-Rank Adaptation (LoRA) technique (Hu et al., 2021a), which efficiently fine-tunes pre-trained models by adapting a small number of model parameters. This method significantly reduces computational requirements while improving performance.

### 3.3 Implementation Details

The experiments were conducted with a learning rate of  $1e-3$ . For fine-tuning the T5-base (Guan et al., 2024) using LORA (Hu et al., 2021a), we selected a rank of 32, a LoRA alpha value of 32, and a LoRA dropout of 0.05. The T-5 small model was trained for 10 epochs while the Flan-T5 model was trained for 2 epochs. All experiments were carried out on a P100 GPU via Kaggle.

## 4 Analysis & Results

The challenge provides two individual datasets of different distributions for training, evaluation, and testing. Details about the distribution of the datasets are illustrated in Table 2. To evaluate the lay summary, we have taken into account three major factors: readability, relevance, and factuality. Readability scores, i.e., Flesch-Kincaid Grade Level (FKGL), Dale-Chall Readability Score (DCRS), Coleman-Liau Index (CLI), and LENS assess how easy summaries are to understand for a lay audience. Relevance scores i.e., ROUGE (1, 2, and L) and BERTScore measure how well summaries capture key points from the original research. Factuality scores i.e., AlignScore, SummaC evaluate how accurately summaries reflect factual information.

The detailed results of our experiments are presented in Table 3. Although none of the models achieved high scores, this can be attributed to the lack of constraints and the limited training and context length of the models. This leads to all models converging to similar scores despite differences in model sizes and training methodologies. However, it is evident that training on both datasets together improves the readability and factuality of the models. This improvement is also observed when fine-tuning a large model using parameter-efficient fine-tuning techniques like LORA (Hu et al., 2021b), which is originally pre-trained on a large text corpus. This might be due to the fact that PEFT (Pu et al., 2023) techniques enable large language models to learn more effectively while retaining previously acquired knowledge. However, this claim for lay summarization would require further experiments to validate.

## 5 Conclusion

We fine-tune and evaluate sequence-to-sequence models like T5 for the task of lay summarization (Challagundla and Peddavenkatagari, 2024) across various training dataset settings and optimization methods such as LoRA. We extended our evaluation beyond traditional accuracy metrics to encompass real-world application considerations like relevance (Montague and Aslam, 2001), readability (Paasche-Orlow et al., 2003), and factuality (Pagnoni et al., 2021). Our analysis revealed interesting trade-offs of fine-tuning pre-trained models using traditional and parameter-efficient methods like LoRA (Hu et al., 2021b). Future work should explore the impact of PEFT-similar models for lay and abstractive summarization.

### Limitations

Due to computational resource limitations, we were able to conduct only a limited number of experiments and were constrained by input token limits. Our access was limited to the Kaggle P100 GPU,

for which we are grateful. This restriction led us to primarily experiment with smaller model sizes, potentially missing the benefits of larger architectures. Our hyperparameter tuning was not extensive, and as a result, our models' performance fell short of high scores on evaluation metrics. This indicates substantial room for improvement. We focused narrowly on sequence-to-sequence models, and these limitations present clear opportunities for future research to build upon and investigate.

## Acknowledgments

We would like to thank the Research Society MIT, an undergraduate interdisciplinary technical society of Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India, for supporting our research.

## References

- Ian L Alberts, Lorenzo Mercolli, Thomas Pyka, George Prenosil, Kuangyu Shi, Axel Rominger, and Ali Afshar-Oromieh. 2023. Large language models (llm) and chatgpt: what will the impact on nuclear medicine be? *European journal of nuclear medicine and molecular imaging*, 50(6):1549–1552.
- Bhavith Chandra Challagundla and Chakradhar Peddavenkatagari. 2024. Neural sequence-to-sequence modeling with attention by leveraging deep learning architectures for enhanced contextual understanding in abstractive text summarization. *arXiv preprint arXiv:2404.08685*.
- Rochana Chaturvedi, Jaspreet Singh Dhani, Anurag Joshi, Ankush Khanna, Neha Tomar, Swagata Duari, Alka Khurana, Vasudha Bhatnagar, et al. 2020. Divide and conquer: From complexity to simplicity for lay summarization. In *Proceedings of the First Workshop on Scholarly Document Processing*, pages 344–355.
- Yongchao Chen, Jacob Arkin, Yang Zhang, Nicholas Roy, and Chuchu Fan. 2023. Autotamp: Autoregressive task and motion planning with llms as translators and checkers. *arXiv preprint arXiv:2306.06531*.
- Chung-Cheng Chiu, Tara N Sainath, Yonghui Wu, Rohit Prabhavalkar, Patrick Nguyen, Zhifeng Chen, Anjuli Kannan, Ron J Weiss, Kanishka Rao, Ekaterina Gonina, et al. 2018. State-of-the-art speech recognition with sequence-to-sequence models. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 4774–4778. IEEE.
- Cagla Colak and Lknur Karadeniz. 2023. **ISIKSumm at BioLaySumm task 1: BART-based summarization system enhanced with bio-entity labels**. In *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, pages 636–640, Toronto, Canada. Association for Computational Linguistics.
- Wang Dandan, Hong He, and Cong Wei. 2023. Wang et al. 2023. cellular and potential molecular mechanisms underlying transovarial transmission.
- Angela Fan, David Grangier, and Michael Auli. 2018. **Controllable abstractive summarization**. pages 45–54.
- Mehmet Firat. 2023. How chat gpt can transform auto-didactic experiences and open education?
- Gaëtanelle Gilquin and Stefan Th. Gries. 2009. **Corpora and experimental methods: A state-of-the-art review**. *Corpus Linguistics and Linguistic Theory*, 5(1):1–26.
- Tomas Goldsack, Carolina Scarton, Matthew Shardlow, and Chenghua Lin. 2024. Overview of the biolaysumm 2024 shared task on the lay summarization of biomedical research articles. In *The 23rd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, Bangkok, Thailand. Association for Computational Linguistics.
- Tomas Goldsack, Zhihao Zhang, Chenghua Lin, and Carolina Scarton. 2022. **Making science simple: Corpora for the lay summarisation of scientific literature**. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10589–10604, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Tomas Goldsack, Zhihao Zhang, Chen Tang, Carolina Scarton, and Chenghua Lin. 2023. **Enhancing biomedical lay summarisation with external knowledge graphs**. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 8016–8032, Singapore. Association for Computational Linguistics.
- Boxu Guan, Xinhua Zhu, and Shangbo Yuan. 2024. A t5-based interpretable reading comprehension model with more accurate evidence training. *Information Processing & Management*, 61(2):103584.
- Yue Guo, Wei Qiu, Yizhong Wang, and Trevor Cohen. 2021. Automated lay language summarization of biomedical scientific reviews. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 160–168.
- Som Gupta and S. K Gupta. 2019. **Abstractive summarization: An overview of the state of the art**. *Expert Systems with Applications*, 121:49–65.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021a. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

- J. Edward Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. 2021b. [Lora: Low-rank adaptation of large language models](#). *ArXiv*, abs/2106.09685.
- Harsha Vardhan Khurdula, Anbilparithi Pagutharivu, and Jin Soung Yoo. 2024. The future of feelings: Leveraging bi-istm, bert with attention, palm v2 & gemini pro for advanced text-based emotion detection. In *SoutheastCon 2024*, pages 275–278. IEEE.
- Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi. 2022. [Text generation from knowledge graphs with graph transformers](#). *Preprint*, arXiv:1904.02342.
- Panagiotis Kouris, Georgios Alexandridis, and Andreas Stafylopatis. 2021. Abstractive text summarization: Enhancing sequence-to-sequence models using word sense disambiguation and semantic content generalization. *Computational Linguistics*, 47(4):813–859.
- Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. 2022. [BioGPT: generative pre-trained transformer for biomedical text generation and mining](#). *Briefings in Bioinformatics*, 23(6):bbac409.
- Gaurav Menghani. 2023. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Computing Surveys*, 55(12):1–37.
- Mark Montague and Javed A Aslam. 2001. Relevance score normalization for metasearch. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 427–433.
- Joel Larocca Neto, Alex A. Freitas, and Celso A. A. Kaestner. 2002. Automatic text summarization using a machine learning approach. In *Advances in Artificial Intelligence*, pages 205–215, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Michael K Paasche-Orlow, Holly A Taylor, and Frederick L Brancati. 2003. Readability standards for informed-consent forms as compared with actual readability. *New England journal of medicine*, 348(8):721–726.
- Artidoro Pagnoni, Vidhisha Balachandran, and Yulia Tsvetkov. 2021. Understanding factuality in abstractive summarization with frank: A benchmark for factuality metrics. *arXiv preprint arXiv:2104.13346*.
- Heiko Paulheim. 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3):489–508.
- A.s. Poornash, Atharva Deshmukh, Archit Sharma, and Sriparna Saha. 2023. [APTSumm at BioLaySumm task 1: Biomedical breakdown, improving readability by relevancy based selection](#). In *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, pages 579–585, Toronto, Canada. Association for Computational Linguistics.
- George Pu, Anirudh Jain, Jihan Yin, and Russell Kaplan. 2023. Empirical analysis of the strengths and weaknesses of peft techniques for llms. *arXiv preprint arXiv:2304.14999*.
- Prajit Ramachandran, Peter J Liu, and Quoc V Le. 2016. Unsupervised pretraining for sequence to sequence learning. *arXiv preprint arXiv:1611.02683*.
- Nicolay Rusnachenko and Huizhi Liang. 2024. nicolay-r at semeval-2024 task 3: Using flan-t5 for reasoning emotion cause in conversations with chain-of-thought on emotion states. *arXiv preprint arXiv:2404.03361*.
- Mohammed Sabry and Anya Belz. 2023. Peft-ref: A modular reference architecture and typology for parameter-efficient finetuning techniques. *arXiv preprint arXiv:2304.12410*.
- Madoune R Seye, Bassirou Ngom, Bamba Gueye, and Moussa Diallo. 2018. A study of lora coverage: Range evaluation and channel attenuation model. In *2018 1st International Conference on Smart Cities and Communities (SCCIC)*, pages 1–4. IEEE.
- Anuroop Sriram, Heewoo Jun, Sanjeev Satheesh, and Adam Coates. 2017. Cold fusion: Training seq2seq models together with language models. *arXiv preprint arXiv:1708.06426*.
- Michael Stubbs. 2004. Language corpora. *The handbook of applied linguistics*, pages 106–132.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.
- Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Zhibo Sun, and Yue Zhang. 2024. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing*, page 100211.
- Qihao Zhu and Jianxi Luo. 2022. Generative pre-trained transformer for design concept generation: an exploration. *Proceedings of the design society*, 2:1825–1834.