CHATDESIGN: BOOTSTRAPPING GENERATIVE FLOOR PLAN DESIGN WITH PRE-TRAINED LARGE LANGUAGE MODELS

JINMIN LI[1], YILU LUO[2], SHUAI LU[3,] JINGYUN ZHANG[4,] JUN WANG[5,] RIZEN GUO[6] and SHAOMING WANG[7]
[1,2,3]Shenzhen International Graduate School, Tsinghua University
[1,4,5,6,7]Wechat Lab 33, Tencent
[1]ljm22@mails.tsinghua.edu.cn , 0009-0000-7098-2589
[2]luo-yl22@mails.tsinghua.edu.cn, 0009-0005-5448-9555
[3]shuai.lu@sz.tsinghua.edu.cn, 0000-0003-3772-1613
[4]zhang304973926@gmail.com
[5]earljwang@gmail.com
[6]rizenguo@tencent.com
[7]mangosmwang@tencent.com

Abstract. Large language models (LLMs) have achieved remarkable success in various domains, revolutionizing tasks such as language translation, text generation, and question-answering. However, generating floor plan designs poses a unique challenge that demands the fulfilment of intricate spatial and relational constraints. In this paper, we propose ChatDesign, an innovative approach that leverages the power of pre-trained LLMs to generate floor plan designs from natural language descriptions, while incorporating iterative modifications based on user interaction. By processing user input text through a pre-trained LLM and utilizing a decoder, we can generate regression parameters and floor plans that are precisely tailored to satisfy the specific needs of the user. Our approach incorporates an iterative refinement process, optimizing the model output by considering the input text and previous results. Throughout these interactions, we employ many strategic techniques to ensure the generated design images align precisely with the user's requirements. The proposed approach is extensively evaluated through rigorous experiments, including user studies, demonstrating its feasibility and efficacy. The empirical results consistently demonstrate the superiority of our method over existing approaches, showcasing its ability to generate floor plans that rival those created by human designer. Our code will be available at https://github.com/THU-Kingmin/ChatDesign.

Keywords. floor plan generation, large language models, user interactions, automatic design, deep learning, pre-train models

1. Introduction

Large language models (LLMs) have made significant strides in various domains, revolutionizing tasks such as language translation, text generation, and question-answering. These models, trained on vast amounts of textual data, have demonstrated the ability to comprehend and generate human-like language. In this paper, we propose ChatDesign, an innovative approach that harnesses the power of pre-trained LLMs to generate floor plan designs from natural language descriptions.

Generating floor plan designs presents a unique challenge that requires the fulfilment of intricate spatial and relational constraints. While existing methods have made progress in considering specific constraints such as room types, adjacencies, and boundaries, they often rely on predefined templates or manual parsing techniques. These approaches limit flexibility and adaptability, hindering the generation of floor plans that precisely meet the user's needs.

To address these limitations, we introduce ChatDesign, which incorporates iterative modifications based on user interaction. Our approach involves processing user input text through a pre-trained LLM and utilizing a decoder to generate regression parameters and floor plans tailored to the user's specific requirements. We employ an iterative refinement process that optimizes the model output by considering the input text and previous results, ensuring the generated design images align precisely with the user's needs.

Notable models like CogView (Ding et al., 2021) and Imagen (Saharia et al., 2022) have showcased the potential of pre-trained LLMs in transforming input data into meaningful representations, pushing the boundaries of AI-powered generation tasks. Building upon this progress, our proposed ChatDesign approach harnesses the power of pre-trained LLMs to generate floor plan designs. By doing so, we address the limitations of existing methods and highlight the potential of LLMs in automating and enhancing the floor plan design process.

In summary, our main contributions are as follows:

- We propose a novel method for the floor plan generation task, utilizing a pre-trained large language model to iteratively optimize the output.
- We introduce an innovative scheme for interactive floor plan generation, empowering users to interactively modify the generated floor plans in line with their changing needs.
- We implement a variety of interaction modalities, including text input, audio instructions, mouse clicks, and drag-and-drop operations.
- The empirical results consistently demonstrate the superiority of our method over existing approaches, showcasing its ability to generate floor plans that rival those created by human designer.

2. Related Work

2.1. FLOOR PLAN GENERATION

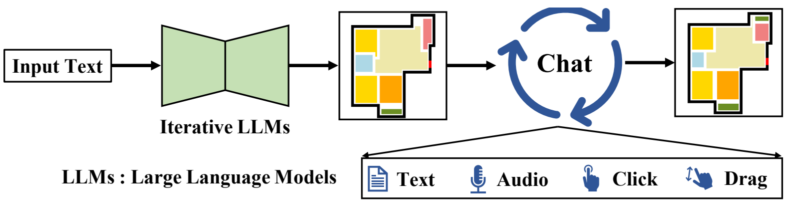
Several methods have been proposed for automatic floor plan design, with most of them taking into account specific constraints such as room types, adjacencies, and boundaries. For instance, Wu et al. (2019) introduced a CNN-based approach that utilizes boundary images as a constraint to determine the location of different rooms. Chen et al. (2020) provided a method where a small set of template-based artificial verbal commands is manually parsed into scene graphs to guide the generation process. Leng et al. (2023) contributed a new dataset called Tell2Design, which consists of over 80,000 floor plan designs accompanied by natural language instructions. Additionally, a benchmark model has been developed to evaluate the performance of different techniques in this field.

2.2. PRE-TRAIN LARGE MODELS

Pre-trained large models (Zhang et al., 2021, Liu et al., 2021) have made significant advancements in various domains. CogView (Ding et al., 2021) utilizes a pre-trained VQ-VAE to transform a target image into a sequence of image tokens. These image tokens are then combined with text tokens and input into a Transformer decoder to generate an image. Imagen (Saharia et al., 2022), on the other hand, is an advanced text-to-image generation model that leverages a large language model T5 (Raffel et al., 2020) for text comprehension and a diffusion model for generating high-fidelity images. Furthermore, Tell2Design (Leng et al., 2023) introduces a new dataset called T2D, which consists of over 80,000 floor plan designs paired with natural language instructions. This dataset enables the development of text-to-floor plan generation models using the powerful T5 language model, known for its exceptional language understanding capabilities. These models have demonstrated impressive results in their respective domains, showcasing the potential of pre-trained large models in pushing the boundaries of AI-powered generation tasks.

3. Methodology

3.1. OVERVIEW



Input Text : *balcony 1 is in south side of the house, next to master room, the size is 32 sqft. balcony 2 is in north east corner of the house, next to kitchen, the size is 32 sqft. bathroom is in west side of the house, next to common room 2 and living room, the size is 48 sqft. common room 1 is in south west corner of the house, next to master room and bathroom, the size is 140 sqft. common room 2 is in north west corner of the house, next to living room and bathroom, the size is 120 sqft. kitchen is in north east corner of the house, next to living room and balcony, the size is 64 sqft. living room is in north side of the house, next to common room 2, bathroom and kitchen room, the size is 220 sqft. master room is in south side of the house, next to common room 1 and balcony, the size is 120 sqft.*

Figure 1. An overview of our proposed ChatDesign.

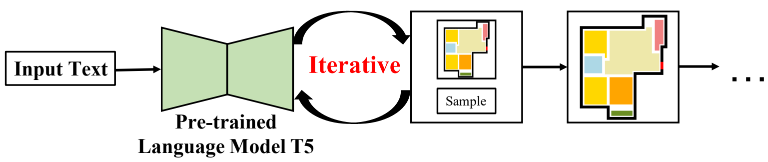
As shown in Figure 1, the methodology of ChatDesign involves an iterative process that leverages pre-trained LLMs to generate floor plan designs based on user input. The overall process can be summarized as follows:

- Input Text and Initial Floor Plan Generation: The user provides a human-described textual input describing their floor plan requirements. This input is processed by the Iterative LLMs, which generate an initial floor plan as a preliminary design. The Iterative LLMs utilize the power of pre-trained language models to comprehend and transform the textual input into a visual representation.
- Interactive Design Phase: Building upon the initial floor plan, the user enters an interactive design phase where they can further refine the design based on their evolving requirements. Various interactive modalities such as text input, voice commands, mouse clicks, and drag-and-drop interactions are supported. These interactions enable specific actions such as precise positioning of rooms, moving them within the layout, resizing rooms according to user preferences, as well as adding or removing rooms.
- Template-Based Interactions: In addition to the textual input, users also have the option to skip the initial step and directly select a template floor plan. Based on this template, users can propose new interactive design requirements, allowing them to modify and adapt the template to suit their specific needs.

The iterative nature of ChatDesign allows users to progressively refine the generated floor plans based on their evolving preferences and requirements. By incorporating interactive design features and providing flexibility in both initial input and template-based approaches, the methodology empowers users to actively participate in the floor plan generation process and achieve designs that align closely with their vision. The entire model is trained in a way and the objective function of the training follows the settings of baseline Tell2Design (Leng et al., 2023).

By doing so, ChatDesign enhances user engagement and ensures personalized output. It transforms the traditional design process into a dynamic interaction, understanding user preferences and delivering tailored results. This user-centric approach facilitates continuous improvement of the model, increasing its efficiency and accuracy over time.

3.2. ITERATIVE GENERATION



Input Text : *balcony 1 is in south side of the house, next to master room, the size is 32 sqft. balcony 2 is in north east corner of the house, next to kitchen, the size is 32 sqft. bathroom is in west side of the house, next to common room 2 and living room, the size is 48 sqft. common room 1 is in south west corner of the house, next to master room and bathroom, the size is 140 sqft. common room 2 is in north west corner of the house, next to living room and bathroom, the size is 120 sqft. kitchen is in north east corner of the house, next to living room and balcony, the size is 64 sqft. living room is in north side of the house, next to common room 2, bathroom and kitchen room, the size is 220 sqft. master room is in south side of the house, next to common room 1 and balcony, the size is 120 sqft.*

Sample1: *bathroom is in west side of the house, next to common room 2 and living room, the size is 48 sqft. common room 1 is in south west corner of the house, next to master room and bathroom, the size is 140 sqft.*

Sample2: *kitchen is in north east corner of the house, next to living room and balcony, the size is 64 sqft. living room is in north side of the house, next to common room 2, bathroom and kitchen room, the size is 220 sqft.*

… …

Figure 2. The framework of iterative LLMs

As shown in Figure 2, the iterative generation process in ChatDesign begins with the user's initial textual input, which serves as the foundation for generating an initial floor plan. To enhance the design further, the methodology utilizes random sampling to create modified textual inputs called Samples. These Samples are then used to generate alternative floor plans, and the most favourable ones are chosen based on evaluation metrics. By integrating these selected Samples into the iterative process, the generated floor plans undergo continuous refinement and improvement, ensuring a closer match to the user's specific requirements. The iterative approach in ChatDesign allows for continuous optimization, ensuring that the generated floor plans increasingly align with the user's vision. Through multiple iterations and refinements, the design is continuously improved based on user feedback and evolving requirements. This iterative process facilitates fine-tuning and adjustments, resulting in floor plans that better meet the user's specific needs and preferences.

3.3. DETAILS

3.3.1. Uniform Instructions Translation (UIT)

Figure 3 illustrates the integration of the Uniform Instructions Translator (UIT) within the ChatDesign system. The UIT is designed to handle various forms of user interactions, such as text, voice, mouse clicks, and drag-and-drop, by translating them into five core operations: Locate, Move, Resize, Add, and Delete. Each operation is accompanied by specific parameters, such as object type, target location, and dimensions, to fulfil different user requirements.

For example, when a user interacts through text input, the UIT can identify keywords and phrases to determine which core operation should be executed. Similarly, when a user interacts using mouse clicks or drag-and-drop gestures, the UIT can recognize the action and translate it into the corresponding core operation.

By translating different user interactions into these five core operations, the Uniform Instructions Translator provides a unified and standardized way to process user input. This allows for a seamless and intuitive interaction experience, enabling users to actively participate in the design process and accomplish their desired modifications or additions to the floor plan.

Furthermore, the UIT's ability to handle diverse input modalities ensures that ChatDesign can cater to a wide range of users with varying preferences and abilities. This inclusivity enhances the overall user experience and makes the design process more accessible to a broader audience.

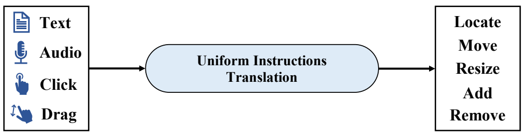


Figure 3. The functionality of Uniform Instructions Translation.

3.3.2. Locate

As shown in Figure 4 (a), the Locate operation is used to position elements within the floor plan. It allows users to specify the exact location or coordinates where a room or object should be placed. This precise control over element placement ensures that users can create floor plans that closely align with their vision and requirements, making the design process more personalized and satisfying.

3.3.3. Move

As shown in Figure 4 (b), the Move operation enables users to relocate rooms or objects within the layout. By providing new coordinates or indicating a target destination, users can easily move elements to desired positions.

3.3.4. Resize

As shown in Figure 4 (c), the Resize operation allows users to adjust the size and dimensions of rooms. Users can specify new dimensions or provide scaling factors to increase or decrease the size of a room. This functionality enables users to fine-tune the floor plan to accommodate specific space requirements, ensuring that the design is both functional and aesthetically pleasing.

3.3.5. Add

As shown in Figure 4 (d), the Add operation permits users to add new rooms or objects to the floor plan. Users can specify the type of room, its dimensions, and any additional details required. This capability allows users to expand and customize their design, adding new elements that cater to their unique needs and preferences, resulting in a more personalized floor plan.

3.3.6. Delete

As shown in Figure 4 (e), the Delete operation enables users to remove existing rooms or objects from the floor plan. By identifying the element to be deleted, users can easily eliminate unwanted components. This feature provides users with the freedom to declutter and simplify their design, ensuring that the final floor plan is efficient, practical, and tailored to their specific needs.

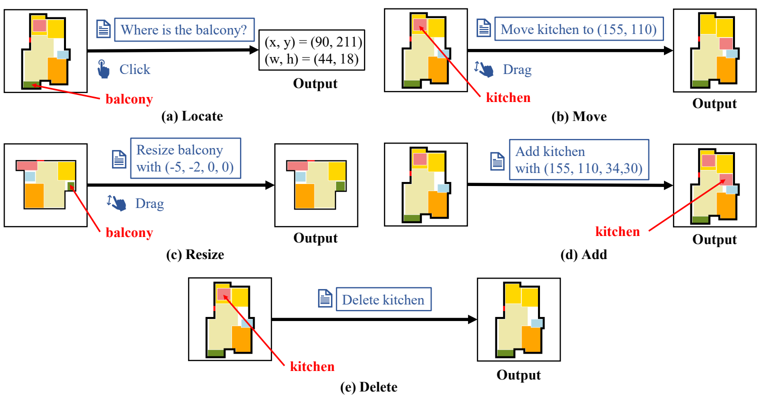


Figure 4. The operations of Locate, Move, Resize, Add and Delete.

In summary, the five core operations - Locate, Move, Resize, Add, and Delete - implemented in the Uniform Instructions Translator (UIT) provide users with a comprehensive toolset for customizing their floor plans. These operations allow users to exert precise control over the placement, size, and composition of their designs, facilitating a highly personalized and user-centric design process. By enabling users to interact with the system in a variety of ways, from text and voice to mouse clicks and drag-and-drop, the UIT ensures a seamless and intuitive user experience. This highlights the strength of ChatDesign in accommodating diverse user preferences and transforming complex design tasks into simple, manageable steps.

4. Experiments

4.1. SETTINGS

4.1.1. Datasets

We utilized the T2D dataset (Leng et al., 2023), which comprises 5,051 manually annotated language instructions and 75,737 AI-generated language instructions. For the T2D dataset, Amazon Mechanical Turk workers were hired and instructed to provide instructions for each room based on the given floor plan image. The requested instructions were expected to reflect the semantic, geometric, and topological information of the floor plan, allowing designers to ideally reproduce the layout based on the instructions. The remaining floor plans were used to generate artificial language instructions by the AI, following predefined templates. Additionally, human designers evaluated and audited the quality of the datasets to ensure their reliability and accuracy.

4.1.2. Evaluation Metrics

To perform evaluations, we employed Intersection over Union (IoU) scores at both macro and micro levels, measuring the overlap between pixel-level ground truth (GT) and generated floor plans. The definitions of these metrics are as follows:

$$\text{Micro IoU} = \frac{\sum_{r=1}^{R} I_r}{\sum_{r=1}^{R} U_r}, \qquad \text{Macro IoU} = \frac{1}{R}\sum_{r=1}^{R} \frac{I_r}{U_r}$$

where I_r and U_r represent the intersection and union of the ground truth and predicted rooms, respectively, for the r-th room type in the floor plan. R denotes the total number of room types. The Macro IoU calculates the average IoU across different room types, while the Micro IoU computes the global IoU by aggregating all rooms.

4.2. MAIN RESULTS

4.2.1. Comparison with state-of-the-art methods

As shown in Table 1, iterative indicates that we employ iterative optimisation with random sampling in the training of the model. Our proposed ChatDesign outperforms all other methods in terms of Micro-IoU. Additionally, ChatDesign-iterative surpasses all other methods in terms of Macro-IoU. These results indicate that our approach achieves superior performance compared to existing methods. Moreover, our methodology approaches the level of performance exhibited by human designers, indicating its effectiveness in generating floor plans that closely align with human expertise and design principles.

As shown in Table 2, we randomly selected 10 challenging examples from the T2D dataset (Leng et al., 2023). The second row represents the performance of Tell2Design, while the third row shows the results after interactive refinement using user feedback. It can be observed that ChatDesign demonstrates an improvement of nearly 10% compared to the original performance. This improvement underscores the value of integrating user feedback into the learning process of AI models, allowing them to better adapt and optimize their performance. Moreover, it highlights the potential of ChatDesign as a powerful tool for tackling complex tasks, demonstrating its capability to learn and improve from interactive refinement.

Method	Micro-IoU	Macro-IoU
Obj-GAN (Li et al., 2019)	10.68	8.44
CogView (Ding et al., 2021)	13.30	11.43
Imagen (Saharia et al., 2022)	12.17	14.96
Tell2Design (Leng et al., 2023)	54.34	53.30
ChatDesign (ours)	**58.31**	55.43
ChatDesign-iterative (ours)	54.57	**57.24**
Human	64.67	62.32

Table 1. IoU between generated floor plans and ground-truth for ChatDeisgn and other baselines.

Method	Micro-IoU	Macro-IoU
Tell2Design (Leng et al., 2023)	47.56	51.10
ChatDesign (ours)	**56.82**	**60.56**

Table 2. IoU between generated floor plans and ground-truth for ChatDeisgn and Tell2Design.

4.2.2. Visualisation Results

As shown in Figure 5, the floor plans generated by our method outperform all the comparative approaches, showcasing exceptional performance. Furthermore, it can be observed that the floor plans generated by ChatDesign are closer to the real-world layout compared to the floor plans hand-drawn by human designers. This suggests that our ChatDesign approach can produce floor plans that exhibit a high degree of realism and accuracy.

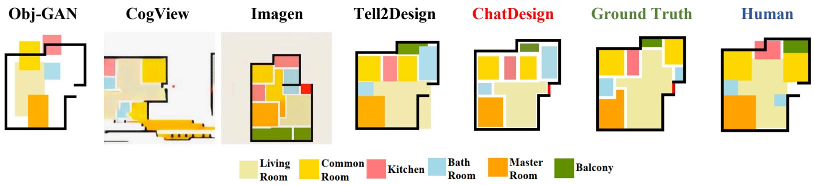


Figure 5. The visualisation results for ChatDesign and other baselines.

Figure 6 indicated that we further optimise the model-generated floor plans using post-processing to further approximate human needs. The main steps include Add door, Erase, Boundary

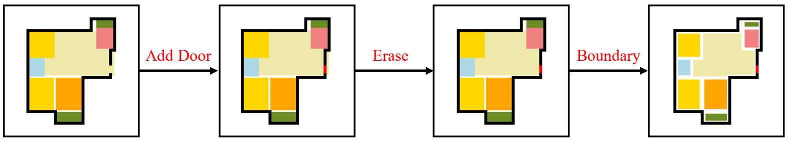


Figure 6. The post-process of model-generated floor plans.

The Add Door step adds entrance doors to the entire space, ensuring accessibility to the floor plan. The Erase step removes unnecessary elements outside the main walls, maintaining the design's integrity. The Boundary step generates boundaries between different rooms, defining individual spaces and enhancing the floor plan's readability and practicality. These post-processing steps refine the floor plan, making it functional and aesthetically pleasing. They contribute to a well-planned, coherent space that meets the user's specific needs and preferences.

4.2.3. Result Analysis

The experimental results demonstrate that our ChatDesign approach performs exceptionally well on the T2D dataset. Compared to other methods, ChatDesign achieves a significant improvement of nearly 10% in terms of performance. The visual results shown in Figure 5 indicate that the floor plans generated by ChatDesign outperform other methods and closely resemble real-world floor layouts.

4.3. DISCUSSION

4.3.1. The Role of the Human Designer

The human designer plays a crucial role in assisting the AI in the floor plan generation process. They provide floor plans as a reference for evaluating ChatDesign's performance, enabling the AI system to learn from their expertise and creativity. The designer also provides subjective ratings for the generated floor plans, serving as feedback for further improvement. This feedback loop allows the AI system to refine its performance and adapt to the designer's preferences. This collaborative approach enhances the abilities of both the human designer and the AI system, resulting in efficient and effective floor plan generation. By combining human creativity and AI automation, the partnership leads to a flexible, adaptive, and user-centric design process, catering to diverse user needs while reducing time and effort.

4.3.2. Future Research

In future work, we suggest exploring the incorporation of additional design constraints, such as building codes and accessibility requirements, to ensure regulatory compliance in the generated floor plans. Furthermore, investigating methods for generating 3D representations would enhance visualization and enable comprehensive design evaluations.

Expanding the training dataset and exploring transfer learning techniques could improve the model's understanding of domain-specific language and enhance its ability to generate diverse and accurate floor plan designs. Conducting user studies with professional architects and designers would provide valuable feedback on the usability and practicality of ChatDesign.

5. Conclusion

In conclusion, this paper introduced ChatDesign, an innovative approach that leverages pre-trained large language models (LLMs) to generate floor plan designs from natural language descriptions. By incorporating iterative modifications based on user interaction, our approach enables precise customization of floor plans to cater to specific user requirements. Through rigorous experiments and user studies, we demonstrated the feasibility and effectiveness of our method, consistently outperforming existing approaches and producing floor plans that rival those created by human designers. Our contributions include a novel method for floor plan generation using pre-trained LLMs, an interactive scheme that empowers users to modify generated floor plans, and the implementation of multiple interaction modalities for a user-centric design process.

Acknowledgements

We gratefully thank the National Science Foundation of China (Grant Nos. 52078294 and 51825802) for funding this work. Jinmin Li and Yilu Luo are co-first authors and Shuai Lu is corresponding author.

References

Leng, S., Zhou, Y., Dupty, M. H., Lee, W. S., Joyce, S. C., & Lu, W. (2023). Tell2Design: A Dataset for Language-Guided Floor Plan Generation. arXiv preprint arXiv:2311.15941.

Wu, W., Fu, X. M., Tang, R., Wang, Y., Qi, Y. H., & Liu, L. (2019). Data-driven interior plan generation for residential buildings. ACM Transactions on Graphics (TOG), 38(6), 1-12.

Li, W., Zhang, P., Zhang, L., Huang, Q., He, X., Lyu, S., & Gao, J. (2019). Ob ject-driven text-to-image synthesis via adversarial training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12174-12182).

Ding, M., Yang, Z., Hong, W., Zheng, W., Zhou, C., Yin, D., ... & Tang, J. (2021). Cogview: Mastering text-to-image generation via transformers. Advances in Neural Information Processing Systems, 34, 19822-19835.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., ... & No rouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding. Advances in Neural Information Processing Systems, 35, 36479-36494.

Chen, Q., Wu, Q., Tang, R., Wang, Y., Wang, S., & Tan, M. (2020). Intelligent. home 3d: Automatic 3d-house design from linguistic descriptions only. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12625-12634).

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research, 21(1), 5485-5551.

Liu, X., Yu, H., Zhang, H., Xu, Y., Lei, X., Lai, H., ... & Tang, J. (2023). Agentbench: Evaluating llms as agents. arXiv preprint arXiv:2308.03688.

Zhang, Y., Yang, J., Yuan, Y., & Yao, A. C. C. (2023). Cumulative reasoning with large language models. arXiv preprint arXiv:2308.04371.