

Application of Generative Artificial Intelligence to Visual Consistency in Product Packaging Design

Qilin Yu*

Ganxi Vocational College of Science and Technology, Xinyu 338000, China

*Corresponding email: 18607902274@163.com

Abstract

Generative artificial intelligence (GenAI) enables rapid packaging concept exploration, but consistent design across product variants remains difficult to specify and verify. This article develops a constraint-based workflow that separates invariant brand elements, variant-specific imagery and colors, and production constraints. The workflow was illustrated with a three-product tea series (Jasmine, Citrus, Berry) and examined in a six-image exploratory generation exercise using Adobe Firefly. Each product was generated once using a narrative prompt and once using a structured-field prompt carrying the same design information. Paired requests specified identical seeds, a 3:4 aspect ratio, and 1MP resolution. A single-observer visual inspection found the intended ingredient motifs, green ornamental frames, empty brand zones, and distinct accent bands were broadly represented in all six images. Unrequested pseudotext appeared in all three structured-prompt images but in none of the three narrative-prompt images. Proportions and motif scale also varied across the family. This small descriptive study does not establish the general superiority of either prompt format, but it demonstrates a traceable approach to packaging design constraints and identifies reasons that artificial intelligence (AI) concepts require human typesetting and technical finishing before production.

Keywords

Generative artificial intelligence, Product packaging design, Visual consistency, Design constraints

Introduction

Packaging serves as both a functional product interface and a visible expression of brand identity. A package must distinguish its product, communicate essential information, and remain recognizable as part of a broader product family. This balance becomes challenging when a brand offers multiple flavors, sizes, or product categories. Changes intended for product differentiation - such as color accents and illustrations of ingredients - can inadvertently disrupt the logo placement, typographic hierarchy, or other elements that customers associate with the brand.

Text-to-image generative AI has expanded opportunities for rapid visual exploration. Diffusion-based image synthesis can transform written descriptions into diverse concepts. While more recent design research indicates that the wording and structure of prompts influence how designers navigate feasible, novel, and aesthetic alternatives [1,2]. These capabilities are pertinent to packaging design because teams often need to generate numerous options before selecting a final concept.

However, a collection of appealing images does not necessarily constitute a coherent brand system.

Consequently, the practical challenge extends beyond image quality. The designer must specify which features are non-negotiable across the entire product family, and which may vary between individual product variants. A generic prompt such as “design a premium tea package” contains almost no information about these critical relationships. Furthermore, repeatedly rewriting prompts risks introducing subtle inconsistencies in shared visual elements. Research on AI in graphic design identifies both valuable creative opportunities and persistent challenges in seamless workflow integration, while packaging-specific studies suggest that AI tools benefit from explicit requirements, systematic evaluation, and thoughtful human-guided refinement [3,4].

This article addresses a focused question: How can design constraints be organized so that generative AI supports the exploration of multiple packaging variants without compromising a consistent brand framework?

Rather than proposing a new model or claiming a measurable performance gain, this paper presents a practical method for expressing, transferring, and verifying design constraints. It employs an illustrative three-variant design brief, a structured prompting procedure, and a visual consistency checklist to demonstrate what a small design team can document both before and after image generation.

The contribution is modest yet operational: a clear separation between shared and variable design elements, a five-stage workflow, and a traceable evaluation procedure. It is particularly relevant to early-stage packaging visualization, where design alternatives must be easily comparable before final brand artwork, mandatory legal text, and technical print files are completed.

Technical background and related research

Text-to-image generation and design control

Latent diffusion models provide a technical basis for image synthesis from textual descriptions by combining latent-space modeling with conditioning mechanisms. For designers, this enables rapid generation of colors, forms, textures, and illustrative motifs. However, text prompting alone does not guarantee stable object placement across separately generated images. When the objective is a collection of related packages rather than one isolated illustration, the stability of shared visual anchors becomes a central design requirement.

Methods such as ControlNet demonstrate that diffusion models can incorporate explicit spatial conditions, including edges and structural guidance [5]. Prompt-to-Prompt editing further illustrates how cross-attention can be used to alter selected aspects of an image while preserving other content [6]. These studies establish technical possibilities for controlled generation and editing, but they should not be confused with capabilities automatically available in every commercial text-to-image tool. A practical workflow must describe the controls actually used and avoid treating a prompt phrase as equivalent to a guaranteed spatial constraint.

Brand identity across a product family

Brand consistency is commonly expressed through recurring marks, color relationships, shapes, and composition rules. Gong discussed the value of coordinated graphic design for brand consistency across

a digital service interface, while research on AI-based logo generation examines visual coherence across tool outputs [7,8]. Although logos, service interfaces, and packages serve different functions, these studies are informative because consistency must be observed across several related artifacts. Packaging adds a further complication: differentiation is required at the same time as continuity.

Research on AI-generated packaging provides a closer application reference. Gao and Li propose an AI-assisted tea packaging design method that combines consumer-need analysis, prompt generation, option selection, and digital refinement. Their study demonstrates a more elaborate decision-making pipeline than the one intended here. The present article focuses instead on the minimum usable specification for a small product family: What should remain fixed, what may vary, and how deviations should be detected. The method deliberately avoids requiring custom model training, market surveys, or multi-criteria optimization software.

Prompt structure and human review

Chong et al. showed that prompt strategies in product design are related to different design exploration goals [9]. This supports treating prompt composition as a design operation rather than an incidental technical instruction. In a packaging family, the prompt should transmit a stable set of shared constraints and an explicit set of variant changes. Nevertheless, an organized prompt does not establish that the resulting image will comply with those constraints. The final package needs a separate review stage.

Review is especially important for package text. Chen et al. identified accurate text rendering as a limitation in conventional diffusion image generation and proposed a dedicated approach to improve it [10]. Accordingly, this article separates generated decorative imagery from production typography. Brand names, ingredients, mandatory notices, barcodes, and legally required statements should be typeset and verified in standard graphics or packaging-layout software rather than accepted uncritically from a generated image.

Materials and method

Research type and material

This study combines a practice-oriented workflow specification with six exploratory text-to-image

generations for a fictitious premium tea family with three variants. It is not a randomized experiment, a systematic literature review, or a consumer preference test. Three visual concepts correspond to Jasmine, Citrus, and Berry flavors. For each concept, both a semantically matched narrative prompt and a structured-field prompt were used. One result was retained from each of the six requested generations, with no exclusions made based on visual success. An author-conducted visual inspection was then performed on the visible image previews.

The case requires a common portrait package format featuring a warm-ivory background, fine dark-green ornaments, a reserved upper-center brand zone, a centered product-specific illustration, and a lower accent band. The documented materials now include a family constraint matrix, a reusable prompt structure, three variant specifications, six provider-hosted generation uniform resource locators (URLs) (containing full prompts and requested settings), an exploratory visual audit, and technical layout schematics. The schematics presented in Figure 2 were drawn manually and are not model outputs.

Definition of design constraints

Design requirements are divided into three classes. First, invariant constraints define brand features that must be repeated: package proportions, background treatment, brand-mark zone, basic color palette, visual hierarchy, and illustrative style. Second, variable constraints specify permissible differences among products: an accent color, an ingredient motif, and a variant label. Third, production constraints ensure the technical viability of the concepts: a reserved typography region, adequate contrast, a printable layout, and space for legally required information.

This separation is important because “consistency” does not imply identical appearance. A product family should enable customers to identify each variant without confusing it for a different brand. Consequently, the workflow does not seek to minimize every visual difference; instead, it preserves selected anchor elements while making selected changes intentional.

Constraint-based AI design workflow

Figure 1 shows a constraint-based workflow schematic for exploratory AI-assisted packaging concept development, which yields an editable family of packaging concepts.

Stage 1: Converting the brief into design tokens

The process begins with a short design brief describing the product, package type, target context, and family membership. Ambiguous phrases such as “modern style” are translated into inspectable tokens. For example, a warm off-white surface, dark-green line graphics, and a top-centered brand zone are easier to track across variants than a broad requirement for “a natural premium look”. Each token should have a clear function: identity anchor, variant marker, or production requirement.

Stage 2: Building shared and variant prompts

A reusable prompt is assembled from a shared block and a variant block. The shared block specifies the overall format, repeated style, main color family, and elements that must remain fixed. The variant block supplies the stock keeping unit (SKU) -specific accent and ingredient motif. The same prompt architecture is used across all products, only the variable fields change. To support a future paired comparison of free-form and structured prompts, both forms should contain equivalent design information so that any difference cannot be attributed merely to one form having more requirements.

Stage 3: Generating and retaining alternatives

Six outputs were generated through Adobe Firefly via the `image_generate` interface on 14th October 2026. Each request specified a portrait 3:4 aspect ratio, 1MP output, one candidate, and a user-specified seed (143401 for Jasmine, 143402 for Citrus, and 143403 for Berry), repeated within each narrative/structured pair. Prompt text was kept equivalent in substantive design requirements. The structured version organized those requirements under named fields. The provider did not disclose the exact underlying model build or independently confirm the actual sampling seed in its response. Full prompt strings, requested settings, request identifiers (IDs), and output URLs are preserved in the accompanying case manifest. Native portable network graphics (PNG) downloads were not included in the local archive, the provider images remain viewable through the recorded URLs while available.

Stage 4: Applying a visual quality gate

Generated concepts are checked against the original specification rather than being judged only by visual attractiveness. This gate evaluates whether brand anchors are recognizable, variant signals are unambiguous, layout regions have been preserved, and unwanted text or distorted elements have appeared. A designer may accept,

revise, or reject a candidate, with a short reason recorded for each decision. This creates a simple audit trail without requiring complex statistical modeling.

Stage 5: Completing the design manually

A visually acceptable AI draft remains a concept rather than a production-ready package. Final work includes placing the approved brand mark, using verified fonts,

composing product-specific labels, checking contrast, constructing the printer's dieline, and confirming product information.

The AI-generated image may supply illustrations or layout ideas, but it should not be treated as proof that dimensions, printing tolerances, or regulatory requirements are satisfied.

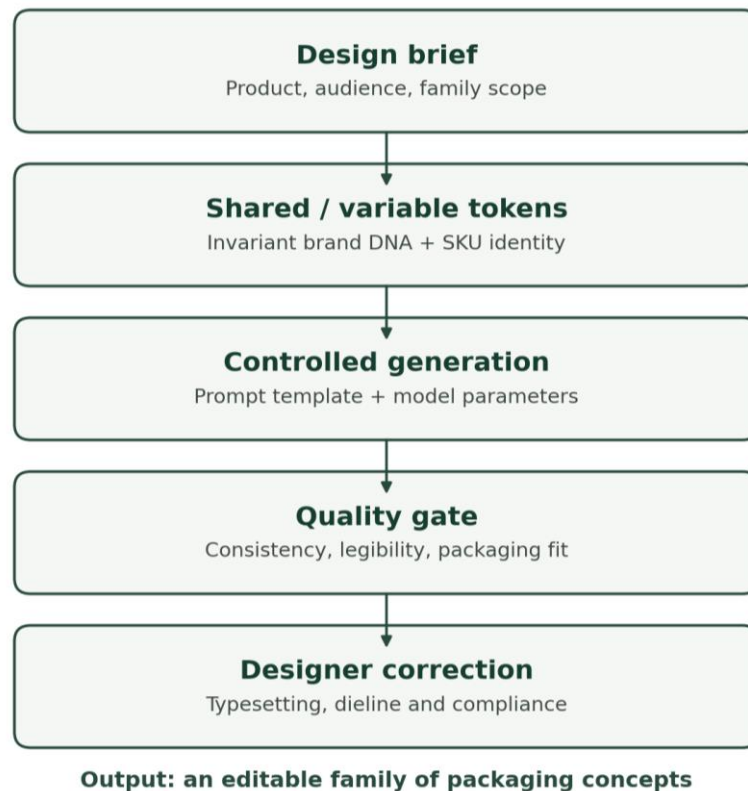


Figure 1. Constraint-based workflow used for exploratory AI-assisted packaging concept development (technical schematic).

Worked example: Three tea packaging variants

Family design specification

The illustrative family employs a vertically oriented front panel featuring a warm off-white field set within a dark-green graphic framework. All three variants reserve the upper center for the brand mark, present an ingredient illustration in the central area, and incorporate a small accent band near the lower edge of the panel. The illustration style is reduced and line-oriented. Precise brand artwork and commercial copy are intentionally omitted from the visual-generation specification, as they require deterministic placement during the finishing stage. The variable tokens are selected to be distinct yet compatible with the common frame. The Jasmine variant utilizes a muted lavender accent accompanied by a floral illustration. The Citrus variant employs a restrained golden accent paired with a sliced-citrus illustration; and

the Berry variant features a deep burgundy accent along with a berry illustration. The palette codes serve as design instructions for the illustrative case and are not measurements extracted from AI-generated images. Table 1 details the practical mapping.

Table 1. Shared and variant design tokens for the illustrative tea packaging family.

Constraint	Design specification
Pack format	Portrait front panel with identical proportions for Jasmine, Citrus, and Berry.
Brand position	Fixed upper-center zone for the final brand mark in all three products.
Base palette	Shared warm ivory #F4F0E8 and dark green #284A3B.
Variant motif	Jasmine flower/citrus slice/berry illustration in one graphic style.
Variant accent	Muted lavender/restrained gold/burgundy in a fixed lower band.
Typography	Reserve empty regions during image generation; verified text added later.

Three visual arrangements

Figure 2 illustrates the target design grammar independently of model generation. The three panels, which are drawn, share a structural frame but differ in their central ingredient motifs and accents.

They remain manually constructed specification

diagrams, not the six experimental images; links to all six provider outputs appear in the accompanying evidence gallery. This distinction is important because a geometrically perfect manual schematic cannot be presented as evidence that an image model met the same alignment requirements.

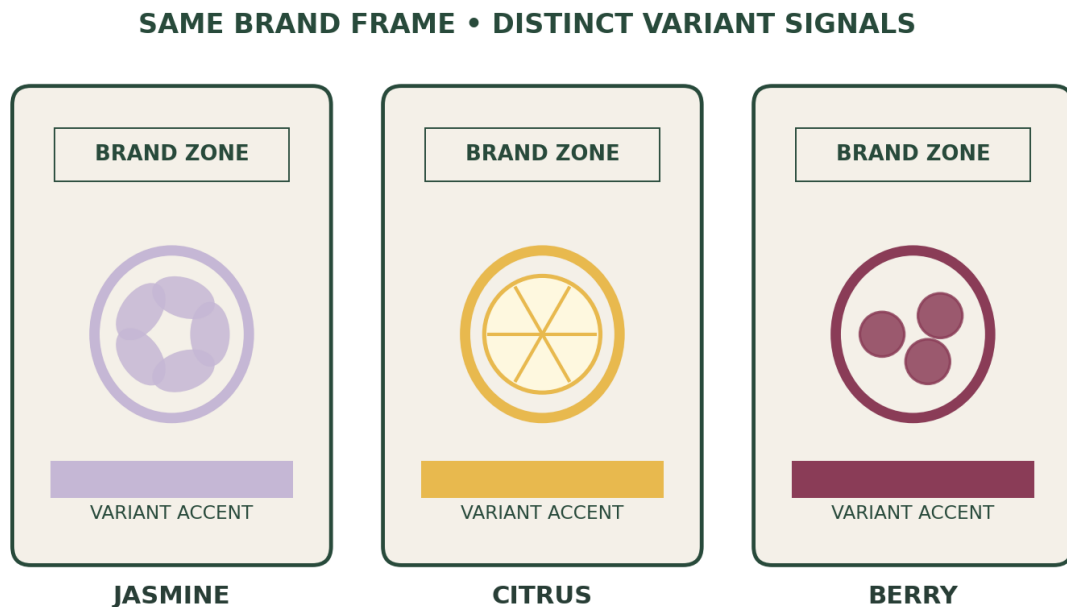


Figure 2. Technical wireframes for a fictitious three-variant tea package family; drawn schematically, not AI-generated outputs.

Prompt organization

For the Jasmine variant, one narrative prompt used was: “Design a premium tea package front in portrait format, with a warm ivory background, a thin dark-green ornamental line border. A fixed empty rectangular brand space at the upper center, a centered flat fine-line illustration of stylized jasmine flowers, and a muted lavender horizontal accent band near the bottom. Leave clear areas for later typography and do not include any lettering or logos. The style and proportions should match a three-product family.” The structured prompt presented the same requirements using labeled fields. The Citrus and Berry variants differed only in the specified ingredient motif, accent color, and paired requested seed. The exact texts are provided in the accompanying manifest.

For the exploratory comparison reported here, the narrative and structured prompt versions preserved identical specifications for the objects, base palette, brand region, main motif, accent hue, empty typography spaces, and the exclusion of lettering and logos. The only difference was the organization of the instructions:

continuous prose versus labeled fields such as “Format”, “Palette”, “Brand Space”, “Central Motif”, “Accent”, and “Exclude”. One generation per product-prompt pairing was retained. This approach controls for the stated instruction content but cannot guarantee matching stochastic sampling or eliminate unobserved service-level changes.

Evaluation procedure and practical implications

A minimal visual audit

The visual quality gate employs the criterion-level checklist presented in Table 2 rather than relying on an invented composite score. An AI-assisted visual review examined the six provider previews for four primary attributes (the ivory packaging field, the dark-green frame, the reserved brand rectangle, and the product-specific motif and accent band) and for any presence of unrequested visible text. This preliminary review is descriptive and requires author confirmation prior to publication. It does not constitute an independent human assessment and was not independently coded, blinded, or assessed using image-level metrics. A criterion was noted

as present if it was recognizable, even in cases where the exact relative dimensions differed across variants.

Table 2. Visual audit criteria applied to the exploratory packaging outputs.

Audit criterion	Inspection question and response
Brand anchors	Are the brand zone, base palette and structural positions retained? If not, revise the prompt or layout.
Variant identity	Are the motif and accent correct and easy to distinguish? If not, adjust the variant tokens.
Visual hierarchy	Is the focal object clear with sufficient space for verified copy? If not, rebalance the artwork.
Typography	Are there false logos, invented words or distorted text? If so, replace with manually typeset copy.
Production fit	Can the design fit a dieline and mandatory label requirements? If not, reconstruct in design software.

Table 3. Results of the single-observer visual audit of six independently generated packaging fronts.

Variant	Prompt	Pseudotext
Jasmine	Narrative	Absent
Jasmine	Structured	Present
Citrus	Narrative	Absent
Citrus	Structured	Present
Berry	Narrative	Absent
Berry	Structured	Present

Presence or absence of unrequested lettering in each preview; one generated image per variant and prompt form. Other specified anchors were broadly recognizable in all six cases, but layout alignment was not measured.

	Narrative	Structured
Jasmine	Absent	Present
Citrus	Absent	Present
Berry	Absent	Present

Unrequested lettering in generated packaging fronts

Figure 3. Observed unrequested lettering in the six generated front-panel concepts (descriptive inspection; n=1 per condition).

Exploratory results and technical interpretation

The six-image application demonstrates that a short set of shared and variant design tokens can produce

recognizable front-panel packaging concepts across three product variants. All six previews showed an ivory-colored packaging area, dark-green ornamental framing, a visually reserved brand rectangle, the instructed botanical/fruit motif, and an accent band of the appropriate hue. This is a presence-level observation, not a measurement of color accuracy or exact geometric alignment. Brand-box proportions, motif scale, margins, and the overall mockup presentation differed. Therefore, the outputs do not establish registration-ready product-family consistency.

The clearest observed failure was invented typography. In the three structured-prompt outputs, extraneous letter-like markings appeared despite an explicit instruction to leave typographic areas empty. The three narrative-prompt outputs had no comparable visible pseudotext in this inspection (in Table 3 and Figure 3). This 3-versus-0 observation is specific to one output per cell and is not evidence that the structured format is generally inferior. The finding highlights a practical quality gate: Generated ornamental packaging drafts must be checked for misleading text before they are rebuilt with verified artwork and typeset information. Studies of text rendering in diffusion models provide relevant technical context.

Comparison with existing research

Compared with a comprehensive AI packaging pipeline that begins with user-review analysis and multi-criteria design selection, the present work addresses a narrower task. It translates product-family identity into promptable shared anchors and independently inspects six first-pass outputs. Relative to general brand-consistency discussions, it explicitly distinguishes invariant frame properties from SKU-specific motifs and accents. Relative to prompt-exploration studies, it contributes a fully recorded, packaging-specific example and a negative-case observation. It makes no new algorithmic claim.

The method is appropriate for preliminary concept work because it requires few resources: a design brief, a compact constraints table, any suitable image-generation interface, and review by a designer. It may also be transferred to label sets, limited-edition product families, or other visually related branded artifacts, provided the designer adapts the shared anchors and production constraints. The visual grammar should not be treated as

universally optimal, since different product categories have different requirements for shape, legibility, safety warnings, and mandatory labeling.

Practical boundaries

Three boundaries are especially relevant. First, generated package mockups can misrepresent dimensions or materials and therefore need technical reconstruction. Second, diffusion outputs may create pseudotext or distorted word forms, making manual typesetting necessary for brand and product information. Third, a visually unified family does not by itself prove improved market recognition or consumer preference. Such claims would require external evaluation, which is outside the scope of the present design-method article.

This study is deliberately small: There was one output for each of the six conditions, no blinded second evaluator, no controlled timing logs, and no objective test of print alignment, physical feasibility, consumer recognition, or market effects. The requested paired seeds do not prove that the service used equivalent stochastic streams. Consequently, causal language and significance tests would be inappropriate. The next validation step is to repeat multiple seeds per condition, export native provider outputs, and introduce an independent design review; none of these planned actions is credited as completed evidence here.

Conclusion

Generative AI can supply recognizable early-stage packaging concepts when shared brand constraints and variant attributes are explicitly specified. A three-product tea series and six actual Firefly generations illustrate the proposed workflow. All six outputs broadly contained the intended family anchors, but layout variation persisted, and unrequested pseudotext appeared in all three structured-prompt samples in this single-pass review. The results support a practical claim regarding feasibility, transparent auditing, and necessary human correction - not a general performance advantage for either prompt style. Production-quality packaging still requires accurate typesetting, dieline construction, and legal label verification. The method and all prompt-level records are provided so that future studies can replicate and extend the exploratory comparison.

Funding

This work was not supported by any funds.

<https://www.wonford.com/>

Data and Material Availability

The supporting case manifest records six individual Adobe Firefly generations with full prompts, requested settings, provider request IDs, source URLs, and Adobe Creative Cloud native-asset identifiers. The six native-generation records were located in provider-managed storage (format: ffgenimg), but their PNG renditions have not been exported into a self-contained offline archive. Access to the provider account or active rendition links may therefore be required for independent image inspection. The case-level visual audit is a single AI-assisted inspection and has not been independently replicated.

Use of Generative AI

Adobe Firefly was used to generate the six exploratory packaging images. AI tools also assisted manuscript drafting and language editing. These tools are not authors; the author must verify the scientific content, source accuracy, and final submission version.

Acknowledgements

The author would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

- [1] Cao, P., Zhou, F., Song, Q., Yang, L. (2025) Controllable generation with text-to-image diffusion models: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4), 4771-4791.
- [2] Oppenlaender, J. (2024) A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology*, 43(15), 3763-3776.
- [3] Li, H., Xue, T., Zhang, A., Luo, X., Kong, L., Huang, G. (2024) The application and impact of artificial intelligence technology in graphic design: a critical interpretive synthesis. *Heliyon*, 10(21), e40037.
- [4] Gao, Z., Li, Z. (2026) A tea packaging design method based on artificial intelligence generated content. *Journal of Engineering Design*, 37(2), 339-371.
- [5] Qi, J., Xu, W., Zhu, Q., Chu, X., Wang, Y. (2026) Collaborative control in diffusion models for precise image generation: a survey. *Mathematics*,

- 14(15), 2737.
- [6] Yang, F., Yang, S., Butt, M. A., Van De Weijer, J. (2023) Dynamic prompt learning: addressing cross-attention leakage for text-based image editing. *Advances in Neural Information Processing Systems*, 36, 26291-26303.
- [7] Gong, X. (2023) Enhancing brand consistency via integration of emotional graphic design and cloud service interface system. *The Design Journal*, 26(4), 673-684.
- [8] Huang, H., Lee, K. (2026) Development of an integrated branding platform based on generative AI and a comparative usability study between designers and non-designers. *Archives of Design Research*, 39(1), 199-222.
- [9] Chong, L., Lo, I. P., Rayan, J., Dow, S., Ahmed, F., Lykourantzou, I. (2025) Prompting for products: investigating design space exploration strategies for text-to-image generative models. *Design Science*, 11, e2.
- [10] Chen, J., Huang, Y., Lv, T., Cui, L., Chen, Q., Wei, F. (2023) Textdiffuser: diffusion models as text painters. *Advances in Neural Information Processing Systems*, 36, 9353-9387.