# Integrating Generative AI into Creative Workflows: Dealing with Consistency, Scene Control, and Refinement in a Professional Image Generation Case Study

**Rodolfo Ocampo Blanco**
Interactive Media Lab
University of New South Wales
Sydney, NSW, Australia
r.ocampo_blanco@unsw.edu.au

**Oliver Bown**
Interactive Media Lab
University of New South Wales
Sydney, NSW, Australia
o.bown@unsw.edu.au

### Abstract

Generative artificial intelligence (GenAI) systems are increasingly being adopted in creative workflows. While understanding these workflows is crucial for designing effective co-creative tools, detailed accounts of these workflows and how generative AI (GenAI) tools succeed or fail to afford them are limited. In this paper, we describe a case study using GenAI to produce part of the AFR Magazine, published by The Australian Financial Review. We identify three main challenges in using these tools: consistency in subject generation, scene control, and refinement and iteration of outputs. We outline the nature of these problems and their immediate solutions for real-world creative workflows.

## Introduction

Generative artificial intelligence (GenAI) systems dedicated to image generation are increasingly being adopted in creative workflows (Davenport and Mittal 2022; Rafner et al. 2023). A 2024 survey by Adobe found that around a third of creative professionals have used image GenAI tools like Midjourney, Dall-E and Stable Diffusion (Offerman 2024) in their work. However, detailed accounts of how these tools succeed or fail to afford effective creative outputs are limited.

In this paper, we critically describe one such workflow as a case study, carried out in collaboration with The Australian Financial Review (AFR) to produce their 2023 Power Issue partly using generative AI. Though this case study we identify three main challenges and consider how they were solved practically, providing an analysis of the problem space.

Specifically, we generated 'impossible photography' for the 2023 list of most powerful Australians (the Power List) published yearly by the magazine. They intended to create photographs that revealed aspects of the subjects' lives, expressing these in creative ways that would otherwise be implausible or impossible with traditional photography, while also sparking discussion around the ethics and risks of deepfakes.

The process involved working with multiple GenAI tools to achieve the objective, often in response to the limitations of other tools. In retrospect, we highlight three main limitations corresponding to essential affordances of creative tools not satisfactorily supported by generative AI visual tools:


Figure 1: AI-generated cover for the AFR 2023 Power Issue. A fold-out cover was designed, with a real and AI-generated portrait on each side. Readers were invited to guess which one was real.

***Consistency*** involves being able to generate the same subject repeatedly in different scenarios. ***Control*** involves being able to place the subject in a specific scene and with a specific visual style. ***Refinement*** involves being able to iteratively refine image outputs towards a final product. Different tools afford these operations to different degrees, but ultimately we found that as users, an workflow required piecing together different tools to address these affordances.

The contributions of this paper are the following: an overview of a generative AI workflow for impossible photography, identification of three challenges faced within it, and an assessment of how current systems succeed or fail to afford them. Lastly, we analyse future directions that could address these challenges.

## Case study details: generating a magazine issue for the Australian Financial Review

### Overview

The Australian Financial Review produces a yearly Power Issue within the AFR Magazine, which lists the most powerful individuals of the year in Australia. For 2023, they wanted to include AI-generated portraits alongside traditional camera-shot portraits to highlight aspects of the Power Listers' lives and work in ways that would be implausible or impossible through normal photography.

This particular creative practice has been termed "impossible photography", and seeks to create wonder, metaphors and new visual narratives (Toromanoff 2017; Shun-liang Chao 2017; Johansson 2012). This practice differs from deep fakes in the sense that it does not seek to deceive. Rather, the images were clearly labeled as AI-generated.

In our case, the editorial team hoped "to see how far we could push the AI to create magazine-quality portrait photos" while at the same time "spur on business leaders, cultural leaders and politicians to think more urgently about those risks [deep fakes], and what to do about them" (Drummond 2023).

For brevity, our outline of the creative workflow is ordered according to the three issues identified.

### Workflow description and identification of three key operations

**1. Consistency: generating the same subject with high resemblance** We found consistency is one of the key challenges of GenAI image systems. That is: generating the same subject or object across different generations, in different settings and placements. We had the added challenge that the generative systems we used were not able to generate our subjects' faces reliably as they were not widely represented in the training data.

For example, we first attempted to generate our subjects with tools such as MidJourney or DALL-E with poor resemblance results as shown in Figure 2 (Figure 2).

Figure 2: Comparison of Midjourney-generated portrait (Left) and actual portrait (right). Actual portrait credit: David Foote-Auspic/DPS, Creative Commons

Figure 3: Two examples of portraits produced via Dreambooth Stable Diffusion fine-tuning.

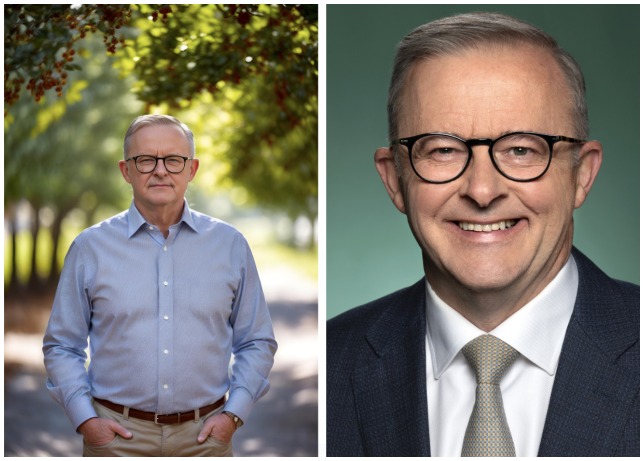

Figure 4: Left: portrait produced by PhotoAI. Right: real portrait

Given this limitation, we turned to explore the training of GenAI models for each Power Lister. First, we collected publicly available images of each subject. We then attempted to train a Stable Diffusion model using the Dreambooth tool. However, the results were inadequate, showing little resemblance and bad quality, as shown in Figure 3.

After testing different tools and techniques, we ultimately chose PhotoAI, a commercial tool that handles the fine-tuning of image models given a small dataset of photos. This tool is normally used to train a model on specific people to generate professional headshots using GenAI. We tested PhotoAI with our subjects (Figure 4), and, determining the photos met the quality standards of the magazine, we decided to use it to generate consistent photos of the Power Listers.

**2. Control: stylistic and structural manipulation** With a trained PhotoAI model for each Power Lister, we attempted to place them in scenes reflecting their personality and work. For example, we wanted to generate a photo of Anthony Albanese DJing, as it is well-known in Australia that the Prime Minister enjoys music and DJing. We used PhotoAI's prompt feature to generate an image of Anthony Albanese with the prompt: "(unique id for Anthony

Figure 5: Generated photos of Anthony Albanese DJing using prompting only. They have high resemblance but are not deemed stylistically or structurally adequate for the magazine issue.

Figure 6: Using an image reference conditioned by depth and photo style yielded better results for Anthony Albanese.

Albanese) DJing at a party with dim and warm lighting". However, the AFR art director considered the generated image's style and scene composition unsuitable for the magazine (Figure 5).

PhotoAI offers image-to-image generation, which allows users to pass an image as a reference and generate a new image that is structurally and stylistically based on the reference to a user-defined degree. This requires a desired reference image, which in itself might be hard to obtain. For this purpose, we used Midjourney, a more powerful generative image system than PhotoAI, to create reference images as inputs to PhotoAI. Midjourney allowed us to generate generic subjects in arbitrary scenarios with greater creative freedom, which were then replaced using PhotoAI with our intended subjects. We did this for each of our subjects, generating or collecting dozens of reference images per subject, and then testing how well our subject was placed in the desired subject. Two examples are shown in (Figures 6 and 7).

**3. Refinement: Iterating Outputs Convergently** The last critical step in our workflow was refining and iterating

Figure 7: Using an image reference conditioned by depth and photo style yielded better results for generating Margot Robbie sitting at the Prime Minister's Office

Figure 8: In-painting usually produced lower quality results, as observed in the hands produced inside the in-painted circled section.

the outputs. Consider the DJ Prime Minister image shown in the left of Figure 8. While it had a high face resemblance and aligned with the structural and stylistic intention, the team felt it needed further improvements specifically, the DJ decks looked empty and the DJ'ing unrealistic.

First, we attempted to use in-painting, a native operation in most generative image AI systems that allows users to erase and regenerate selected areas of an image. However, in-painting often produced lower quality results with structural flaws as shown in Figure 8.

Another attempted approach was to feedback the image as a reference for a new generation. However, we did not successfully achieve the intended edit. As a workaround, the team identified a previous generated image that had the desired DJ deck images, but the facial expression and gaze were suboptimal. The team used Photoshop to combine both halves of each image to produce the final printed version, shown in Figure 9.

We found that iterating outputs was the most challenging step in our workflow using generative AI systems, and this has been highlighted by other practitioners (Novak 2024).

## Analysis

We identified three key limitations for generative AI systems to be adopted in a real-world creative workflows: consistent subject generation, scene control, and refinement and iteration of outputs. These correspond to critical steps in our


Figure 9: The final portrait printed on the cover of the AFR Magazine, newspaper version on September 29th, 2023.

workflow which are not satisfactorily afforded by individual generative tools.

Consistent subject generation was a challenge in our workflow but it also remains a vital challenge in other visual production workflows, and it is often highlighted by practitioners (Dylan 2023; Novak 2024). New techniques, such as Dreambooth, textual inversion, and LORA, offer promising directions for consistent subject generation (Ruiz et al. 2022). However, these techniques can often produce suboptimal results without extensive fine-tuning and engineering, which provides a hurdle for creative practitioners. On the other hand, tools like PhotoAI manage this complexity and offer a simpler interface for training models on specific subjects for consistent generation. But, in the case of PhotoAI it is difficult to steer away from the professional headshot style, given its background prompt engineering and style fine-tuning.

Secondly, beyond generating the same subject consistently, the ability to control the structure of the image, composition, and scene is crucial in visual workflows. There has been some progress in this area, particularly with techniques like image-to-image that allow image-conditioned generation in addition to text-conditioned generation (text-to-image). Out of our three creative challenges, this was the one more readily afforded by current tools. However, is it still inadequate at enabling fine control of an image, supporting practitioners' comparisons to playing with slot machines (Dylan 2023) .

Lastly, refinement is the least afforded requirement among the three. This is closely tied to consistency, as refining requires the subject to remain essentially the same with some details variation, and also tied to control since changes often require some degree of scene control. New techniques like InstructPix2Pix (Brooks, Holynski, and Efros 2022) allow guided text-based edits. However, these techniques are still limited. Moreover, some creative practitioners argue this is the most important part of a visual creative workflow (Novak 2024).

Our case study highlights a lack of creative control through these three shortcomings in affordance, and reveals a workflow in which at least four interactive generative tasks are adaptively managed by the creative user: fine tuning a model to generate accurate images of a person; generating original context images with multiple variations; combining reference image and prompt in a fine-tuned generator; editing together variations. We note specifically that although in-painting is a well known technique that allows targeted refinement of images, it was less effective than manual mixing in of pre-generated context image variations in this professional context.

## Conclusions

From these observations we speculate that this kind of integrated multi-stage workflow is unlikely to be easily superseded by integrated generation tools that remove these individual problems. The text-prompt interface is remarkably powerful at transfer from language to associated image properties, but is entirely inadequate as an interface that provides control over image detail, highlighted in each of these limitations. Instead we expect to see a continued trend for dedicated generative processes to consolidate into an ecosystem of interacting tools.

Lastly, we acknowledge the line between using GenAI tools for our impossible photography workflow and for generating deep fakes is blurred, and this is an important ethical consideration. While deep fakes have the explicit intention of deceiving with malicious intent, our magazine images were clearly flagged as AI-generated. The editorial intent of using GenAI as part of this issue was two-fold: exploring the creative potential of GenAI to tell stories using a new medium of their Power Listers and starting a dialogue around the ethical and social implications of deep fake GenAI technology. Addressing the outlined challenges needs to be balanced with proper mitigations of potential misuse, which is in itself a complex technical, social, policy and interaction design problem.

## References

Brooks, T.; Holynski, A.; and Efros, A. A. 2022. InstructPix2Pix: Learning to follow image editing instructions.

Davenport, T. H., and Mittal, N. 2022. How generative AI is changing creative work. *Harvard Business Review*.

Drummond, M. 2023. What we learnt when making AI images for the 2023 power issue. https://www.afr.com/politics/federal/what-we-learnt-when-making-ai-images-for-the-2023-power-issue-20230830-p5e0jp. Accessed: 2024-3-4.

Dylan. 2023. Stop using text-to-image.this blender + real-time latent consistency workflow is way more fun, and shows how you can use generative AI collaboratively, instead of as a creative slot machinetry it yourself here https://t.co/QyiR5IDQI2 pic.twitter.com/pSsMiA33Rj.

https://twitter.com/dylan_ebert_/status/1724885074313642424 . Accessed: 2024-2-29.

Johansson, E. 2012. Impossible photography — erik johansson.

Novak, M. 2024. Former pixar animator gives one big reason AI video won't work in hollywood. https://gizmodo.com.au/2024/04/former-pixar-animator-gives-one-big-reason-ai-video-wont-work-in-hollywood/. Accessed: 2024-4-30.

Offerman, S. 2024. Creative pros see generative AI as part of their future. https://blog.adobe.com/en/publish/2023/03/21/research-creative-pros-see-generative-ai-as-part-of-their-future. Accessed: 2024-3-13.

Rafner, J.; Beaty, R. E.; Kaufman, J. C.; Lubart, T.; and Sherson, J. 2023. Creativity in the age of generative AI. *Nat Hum Behav* 7(11):1836–1838.

Ruiz, N.; Li, Y.; Jampani, V.; Pritch, Y.; Rubinstein, M.; and Aberman, K. 2022. DreamBooth: Fine tuning Text-to-Image diffusion models for Subject-Driven generation.

Shun-liang Chao. 2017. The alchemy of photography: "grotesque realism" and hybrid nature in jerry uelsmann's photomontages. *Criticism* 59(2):301–328.

Toromanoff, A. 2017. *Impossible Photography: Surreal Pictures that Challenge our Perception*. Hardie Grant Books.