Image-Based AI for Industrial Design

How aligning semantics among designers helps them use AI tools more effectively

Graduation Project IPD
Bouwe Theijse

Chair - Derek Lomas
Coach - Vera van der Burg
Company Coach - Stef de Groot

Graduation Project IPD
Bouwe Theijse
4659651
April, 2024

Chair - Derek Lomas
Coach - Vera van der Burg
Company Coach - Stef de Groot

Cover image was created using a self trained LoRA model and Stable Diffusion 1.5

IMAGE-BASED AI FOR INDUSTRIAL DESIGN

ABSTRACT

This research explores the opportunities and limitations of image-based AI tools for industrial design. In collaboration with designers at Royal Gazelle, several tools were tested. AI tools can broaden and speed up the creative processes but also lack control for the user. AI models and designers use different terms to articulate perceptions and control output. This misalignment stems from misalignment among humans which translated to the training datasets of AI.

This report proposes co-creative image labelling sessions to align perception and articulation of perception in design teams. The aligned perception creates a strong vibe. Collectively labelling an image dataset increased alignment of the perception of images by 30.8% when comparing post- to pre-session questionnaire results. Additionally, participants used more of the same vocabulary after the session.

Training a low-rank adaptation model on these labelled datasets could externalise tacit design knowledge and can be used for further development on AI models that align with designers.

PREFACE

I'm writing the preface two weeks before my final presentation if everything goes according to plan. It has been more than six and half years since I started my Bachelor's in Delft. Studying Industrial Design Engineering has been a blast, but definitely not all the time. My study has taken me to many wonderful places, inside the Netherlands and across the globe. During these years I have met countless wonderful people. I am incredibly grateful for those people I met, who have taught me things, both at uni and outside of it. Thank you!

Over the last few months, I got the amazing opportunity to graduate at Koninklijke Gazelle. This is probably the most "Delft" place to graduate, being the Netherlands' biggest bicycle brand, the ultimate place to graduate as "universitair fietsenmaker". Although only sketched just a few bikes during my graduation. And the only one I actually assembled and repaired was my broken-down Batavus... All jokes aside, I had a great time at Gazelle and thank everyone in the innovation team who was there to welcome me and share their thoughts and enthusiasm for AI. Most of all, I want to thank Stef de Groot for being there throughout the entire journey, coaching me on what went well, but also with points that needed some attention.

My interested in AI for design started during my minor at in Singapore. My professor taught me about tacit design knowledge and I saw a really cool TED talk by Blaise Agüera y Arcas (2016) about image generating AI. Image-based AI tools properly took off in 2022, after which I experimented with several of them during my university and extracurricular projects. I am very glad that I got the time and freedom to dive into the various tools and come up with a unique research direction. Vera van der Burg and Derek Lomas gave me great guidance in this rapidly changing field. Thank you for all the tips, feedback, and sparring sessions!

Last I want to thank my parents and family for their eternal support during my studies and graduation and all my friends who have supported me during my graduation. Whether this was in the form of proofreading the paper, participating in the pilot, sharing coffee and cookies during the occasional lengthy break, asking if I was doing okay on my second official graduation day or simply offering some occasional well-needed distraction. Thank you all for your support!

That is enough "thank you" for now, let's dive into the topic of image-based AI for industrial design!

Figure 0.1. A collection of images created using the LoRA model described in Further Studies.

EXECUTIVE SUMMARY

This report describes the benefits and limitations of image-based AI in the industrial design practice. The research was done in collaboration with designers at Royal Gazelle, the biggest Dutch bicycle manufacturer. After noticing the rapid development of generative and image-based AI, they were curious about how to implement it in their workflow.

Besides the opportunities and limitations of image-based AI for industrial design, this research highlights one specific obstacle: different descriptive terminology between AI and designers causes limited control when interacting with AI. This misalignment stems from inconsistent perception, and articulation of perception, of imagery amongst humans. Individuals look at and describe images from their perspective. Therefore, AI models, trained on human-labelled image datasets, never align perfectly with all users.

To solve this problem, designers could train image-generating AI models based on their personal perception and articulation. Additionally, design teams could co-creatively train models creating and facilitating a shared "vibe" where both parties perceive and articulate the perception of imagery similarly. Ultimately, co-creative labelling should strengthen the vibe within a design team and between designers and the AI they trained.

Survey results indicate that co-creative labelling of an image dataset increases alignment within design teams. After collectively labelling the images, the interpretation of images became 30.8% more similar, participants used more identical terminology, and they mentioned more of the same visual aspects. The designer-trained AI model appears to understand the designers better but partially comprehends the complexity of the designers' articulation of perception. To conclude, collectively training an AI strengthens the vibe between humans themselves and humans and AI.

The report covers the topic of AI for industrial design through four key sections: problem finding, solutions finding, the paper, and future research.

PROBLEM FINDING

The problem-finding section first covers the findings from the design process at Royal Gazelle and then explains the opportunities and limitations of image-based AI for industrial design. It concludes with the opportunities stemming from both research directions combined.

The design process at Gazelle starts when product management decides that Gazelle needs to update or broaden their product portfolio and ends when the designer hands over their 3D design in Solidworks to the engineering department. The product manager investigates new opportunities or product updates and compiles these in a proposition with information like the target audience, competitive environment, etc. The designer takes the proposition and creates a design they initially explore through 2D and 3D sketching. The designer considers numerous aspects like aesthetics, component dimensions, ergonomics, functionality, quality, producibility, and brand consistency to work towards the final 3D design in Solidworks. Each designer from the 3-person design team works on projects individually while complementing the collective brand. Effective communication and a unanimous understanding of the brand design language are essential to an effective design team. Better-aligned teams can design better-aligned and more recognisable product families, eventually leading to more customer acquisition and binding.

Image-based AI tools can broaden and speed up ideation by generating inspiration and idea visualisations. Additionally, AI's bias can be altered because it functions based on editable training data, unlike humans who function based on personal experience. Therefore, it can output ideas different from what users might imagine. However, generative AI has several downsides including authorship of images, risk of tunnel vision, lack of diverse output, and a lack of control. When designers tested the AI tools, their biggest obstacle was formulating their instruction so that the AI would generate output that fit their expectations. AI can only follow instructions if it understands them, hence, the users must be clear and precise, and capture all the essentials in the instructions. This highlights two key challenges: firstly, textually explaining the full complexity of imagery is challenging, and secondly, AI struggles to generate output that fits all features from lengthy detailed prompts. The AI and users have different perceptions and articulation of imagery and therefore communicate ineffectively.

People can train AI models to increase the mutual understanding of images and labels. However, before training an AI model, a design team needs to align within itself, before it can align with and train an AI model. Co-creatively labelling an image dataset forces designers to agree on their perception and articulation of their perception of imagery, i.e. strengthen the "vibe".

SOLUTION FINDING

The solution-finding section explores how to approach co-creative image labelling to effectively align the vibe within a design team and between AI and designers.

Due to AI's data-driven nature, most models require massive amounts of data to function, complicating the training process. For example, Stable Diffusion, one of the most popular text-to-image generators, was trained on 2.3 billion images. Low-rank adaptation (LoRA) models are smaller models that can be trained on as little as 5 images. When there is little time or data to train an AI, LoRA models are ideal. The LoRA model requires a dataset of images with labels - labels are short textual descriptions of the pictures - for training after which the model learns to recognise the patterns in the dataset. To create the dataset, the design team must align their perception and articulation of perception before collectively labelling the images. This strengthens the vibe within the design team. Training the AI model on the team's dataset could create an externalised version of the team's vibe, the future research chapter further explores this direction.

To test the benefits of co-creative image labelling, the design team at Gazelle collectively labelled 12 images of form studies. Form studies are 2D visualisations of 3D shapes in which designers explore how different shapes communicate different aesthetics. The designers first labelled each image individually, afterwards, they formulated a collective label for each image by combining each individual's label. They labelled a total of 12 images.

The quantitative and qualitative analyses indicate that the co-creative image labelling session positively influences the design team's alignment. Design team members interpret images and terminology more similarly, meaning they vibe more closely with each other. They transferred their group vibe to an AI model that better understands their knowledge. All participants reflect that the session significantly aligned their vocabulary to discuss shapes, additionally, all participants felt more aligned on describing images and had a better understanding of how the group would describe shapes.

To conclude, co-creative image labelling of LoRA model training increases the alignment of perception and articulation of perception of imagery amongst designers in design teams.

PAPER

The paper presents the research on the co-creative image labelling session in an academic format with background and related work, preparatory considerations, study setup, findings, discussion, and conclusion.

FUTURE RESEARCH

The last chapter covers two additional experiments: training LoRA models to increase mutual understanding amongst designers in a design team and externalising knowledge; and creating form studies to validate the effectiveness of the alignment within a design team. Lastly, a short section mentions several other interesting directions for future research beyond this graduation project.

Firstly, to validate the effectiveness of the alignment, two designers were tasked to turn a textual description into a form study fitting a given silhouette. Both designers used the same prompt and translated this into visualisations with similar characteristics, however, both shapes are different. This indicates how text, even when aligned, cannot always fully indicate what a picture should look like. This is similar to the obstacle designers faced when using image-generating AI.

Secondly, to validate if training a LoRA model strengthens the vibe between AI and designers, a single designer generated several images with a separate LoRA model. The designer trained this LoRA model on 23 images he labelled individually. He evaluated if the generated images fit his prompts. The subjective terms like "geometric" or "powerful" were caught but the LoRA did not recognise all detail. The LoRA model externalises the designer's vibe but does not capture the full depth. Nevertheless, it indicates a significant increase in mutual understanding between AI and designers.

The last section mentions other directions for future research. Most prominently: repeating the co-creative image labelling sessions in more groups to validate its effectiveness, and exploring the interaction with LoRA models for the transferral of externalised knowledge.

CONTENTS

TERMINOLOGY

Artificial intelligence (AI) is a computer system's approach to simulate human intelligence. A data-driven training allows AI to recognise and recreate patterns. This technology is used in numerous applications, from decision-making to generating images.

Design style consists of the aesthetics embodied by a product. Aesthetics can make the product recognisable as a standalone product or as part of a product family.

Generative AI is AI which generates output like text, images, videos, or music based on the data it was trained on.

Low-rank adaptation (LoRA) models are small AI models that function within bigger models. These models require less training data but can generate high-quality output.

Perception and articulation of perception refers to how people look at things, and how they articulate what they see. When looking at a picture, they may notice different things. If they notice the same things, they may mention things using different words.

A **product family** is a collection of products that matches through their functionality, aesthetic style, or branding. A clear example is the Volkswagen ID series. These electric cars share a similar design style and are all electric.

Product portfolio consists of all products and services offered by a company. For example, Gazelle's product portfolio consists mainly of bicycles, but also offers some accessories and services.

Semantics is the study of meaning in language. In this report, semantic misalignment refers to people attaching different definitions to the same word. This paper focuses on semantic alignment, which refers to bringing the different definitions people attach to words together as closely as possible.

Vibe in this report refers to the perception and articulation of perception. If two team members perceive and articulate perception very similarly, they are vibing, if they perceive and articulate perception very differently, they do not vibe.

PROBLEM-FINDING

The problem-finding chapter highlights three crucial insights: the desire for a unified recognisable brand; the misaligned perception and articulation of perception among designers potentially causes misalignment between designers and AI; and the lack of control when using generative AI tools.

The chapter first goes into the design process at Royal Gazelle and then explains the opportunities and limitations of image-based AI for industrial design. It concludes with the opportunities stemming from both research directions combined.

KEY TAKEAWAYS PRODUCT DEVELOPMENT AND DESIGN PROCESS

1. Designers and product managers want to strengthen Gazelle products' recognisability. However, ambiguous verbal, textual, and visual communication within the design team and reference documents leave too much room for interpretation, complicating this effort.
2. Interaction between colleagues happens mainly in person, most find it easier to discuss topics when they can see each other's body language.
3. Designers experience difficulty explaining their design reasoning to other stakeholders within Gazelle, however, most stakeholders do not necessarily want to hear about the design process which explains the designer's reasoning.
4. All creative work at Gazelle happens digitally. The designers start in Procreate, continue in Gravity Sketch, and finalise their designs in Solidworks. They produce prototypes through 3D printing, possibly with a colour finish.
5. Due to Gazelle's low-risk approach to innovation, their design is more evolutionary than revolutionary. Gazelle design would benefit from a future direction for the Gazelle product portfolio.
6. The designer considers numerous aspects like aesthetics, component dimensions, ergonomics, functionality, quality, producibility, and brand consistency to finalise a 3D design in Solidworks.

KEY TAKEAWAYS AI AND DESIGN

1. Due to a lack of control as captured in the crooked bowtie effect, AI is mainly suitable for inspiration and early ideation, it performs best when there are no strict requirements, constraints, or overly complex ideas.
2. AI alters the developmental exploration of ideation by going directly to high-level output. Therefore, the designer loses friction points where he would normally learn to better understand the problem. Designers must be cautious not to get tunnel vision.
3. AI can speed up and broaden the generative process.
4. AI can be trained on datasets outside a designer's expertise or knowledge and could thus generate images different from what a designer imagines.
5. AI could cause designers to neglect images for the holistic concept even if there are interesting or inspiring parts used in a concept.
6. AI triggers high complexity in ideas but might funnel ideation in mainstream directions due to the training on existing imagery.
7. AI does not always understand the reasoning behind design decisions but it can quickly generate a lot of output to attempt many possible solutions.
8. AI works best when the user combines multiple tools and uses each one for a subtask that they perform well at.

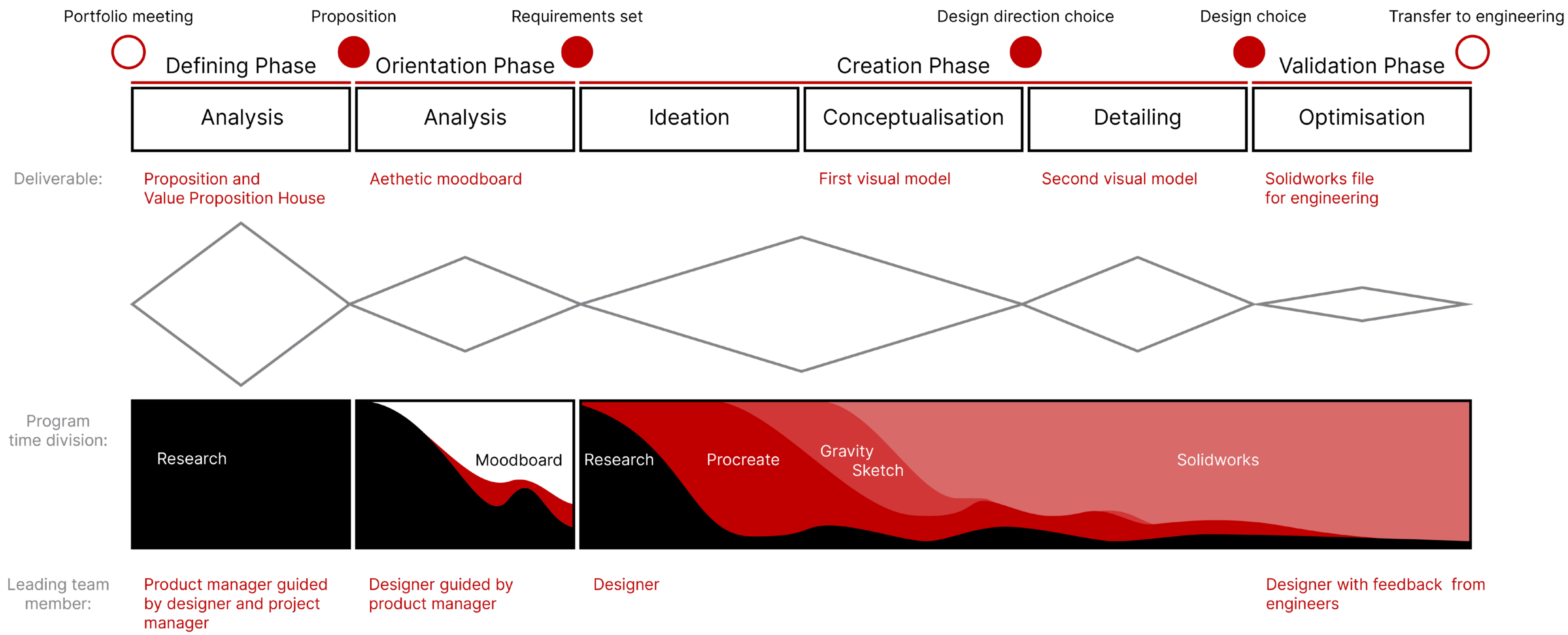


Figure 1.1. A schematic overview of the design process showing the defining, orientation, creation, and validation phase with the deliverables. Additionally it shows the tools used, and the person leading the process per phase.

THE PRODUCT DEVELOPMENT PROCESS

This section goes into the product development and mainly the design process at Royal Gazelle. It mentions how the process starts, progresses, and ends, who is involved, and what the goals are. Figure 1.1 summarises the design process at Royal Gazelle.

PROJECT TEAMS

Gazelle works with project teams consisting of a product manager, a project manager, mechanical engineers, a quality engineer, a manufacturing engineer, a procurement specialist, and a data manager. The product manager initiates new projects after recognising a need to update or expand Gazelle's product portfolio. Figure 1.1 illustrates the product development process from the project initiation until the design gets handed over to the mechanical engineers. The design process consists of the orientation, creation, and validation phases.

DEFINING PHASE

Gazelle evaluates their portfolio during portfolio meetings. If they find the need to update or expand their portfolio, product managers research if a project is worth pursuing. During the defining phase, they research the project goal and summarise key information in a proposition which theyh gives to the designers. That is when the design process starts.

THE DESIGN PROCESS

Designers go through three key phases when designing a product at Gazelle: the orientation, creation, and validation phase.

After receiving the proposition, designers initiate the orientation phase. The proposition forms a basis for designers to understand the project scope. It is the designers' task to interpret the product manager's proposal and create a design that fits the company and customer needs based on the targeted market segment, trends, and target group. Therefore, designers usually maps the aesthetics fitting the target group.

During the orientation phase, designers validate their assumption and gains extra knowledge. Designers perform competitor and trend analysis to better understand the market and customer. For example, they collect retailer consumer insights through questionnaires and interviews. Additionally, they dig into the technological possibilities and components to use. Essentially, they aim to understand the customer's needs and wants; their expectations, understandings, and possible product improvement points. To visualise their findings, designers create mood boards, draw textual conclusions, and collect these in Miro (Miro, n.d.). After the research, designers set the requirements and go into the creation phase.

If the project already has a clear goal, for example updating a previous design, the defining and orientation phase can be sped up or skipped altogether.

During the creation phase, designers do additional research but spend most of their time creating, visualising, and modelling ideas. They start with 2D line sketches, usually from the side perspective, and then moves to sketches from a more informative perspective. They sketch digitally in Procreate (Procreate, n.d.). Depending on a project's timeline and complexity, they move into 3D modelling with Gravity Sketch to get a better understanding of the product's 3D shape (Gravity Sketch, n.d.). They finalise the design in Solidworks to focus on the final detailing and the exact measurements of components (SolidWorks, n.d.). To validate the design direction, designers present a full-scale physical 3D-printed model. After confirming the design direction choice, the designers re-iterate the final design and deliver the final proposal for the design choice which concludes the creation phase.

During the verification phase, the designers test the 3D-printed mock-up. Together with the team, they validate design and customer-focused features. The tests also validate the probable use and misuse of product features. Additionally, the function requirements get validated. Apart from tests, the designer uses the model to present the chosen concept to the directorate.

After the directorate approves, the design gets transferred to the mechanical engineering department which prepares it for production and quality checks. During this process, the designers offer design support to check if changes made by the mechanical engineers fit within the envisioned design form.

BRAND IDENTITY AND AESTHETIC COMMUNICATION

At Gazelle, designers and product managers have both expressed their desire to create a stronger brand identity. This lies both in the design and marketing of the brand. They collaborated with a product and brand design agency to set their brand goals and specific brand identity. During a session with the agency, a team from Gazelle was tasked to rank multiple images on a set of 4 axes in two teams of three. Each axis had opposing labels at opposite ends, for example, one axis had the labels organic and geometric. The session goal was to determine where on these axes Gazelle currently placed itself, and where it wanted to be. They want a strong brand identity to implement over their product portfolio and marketing to become more recognisable and increase customer binding.

While the results of the session were useful, the placement of the different pictures when comparing between groups was especially interesting. Some images were placed on the same axis by both teams but at different ends as shown in Figure 1.2. This hints towards different perceptions of the images which is problematic for designers communicating aesthetic directions.

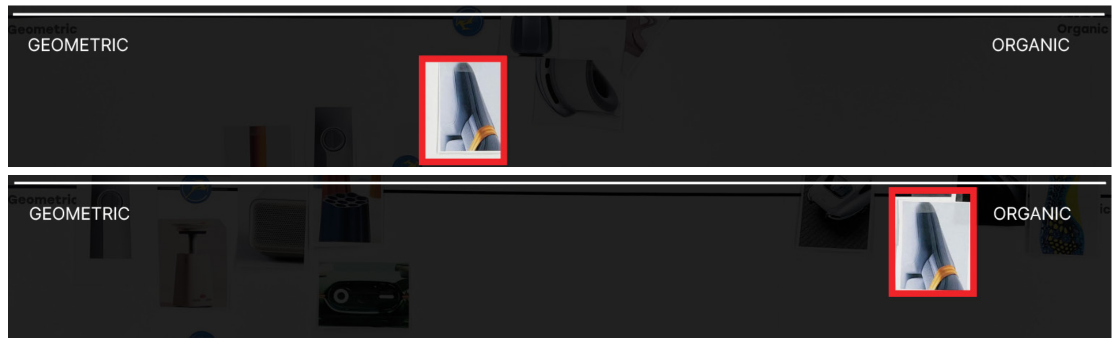


Figure 1.2. Two groups rated the same image differently on the same axis going from geometric to organic.

When designers are developing their idea, they get an idea of the visual style they want to use for a product. Ideally, this fits the style of Gazelle so the new design strengthens the brand identity. Both Gazelle's design style and the envisioned design styles for new products are communicated using mood boards - a collection of aesthetically communicative images - however, the pictures on the mood boards prove to be open to interpretation due to personal perceptions.

I was under the impression that within a company, the collaborative process naturally aligned aesthetic perception and articulation. I assumed that a design team learns to perceive images similarly by valuing the same visual aspects. Additionally, I assumed that the designers would use the same terminology to articulate their perceptions. However, when discussing the perception and articulation of perception with a designer at Gazelle, he mentioned that each team member gave a slightly different twist to the Gazelle design style. Each designer interprets the mood board capturing the Gazelle design style slightly differently which causes the translation of the design style to the product.

The alignment of perception and articulation of perception caused greater mutual understanding in the design team and helped the designers to think and communicate more similarly. This helped them create a shared vibe in which they effortlessly understand each other through the same perception and vocabulary. Strengthening the vibe within the design team would facilitate aesthetically aligning Gazelle's design portfolio.

Creating a shared vibe within the design team is especially valuable considering the evolutionary innovation process at Gazelle. They develop based on the previous version of their bikes to maintain their brand identity. Alternatively, they could set a clear design goal of where they want to be in 10 years. They have done this before by designing a concept bike (Italdesign, 2017). The concept bike set an aesthetic direction, however, they did not design a new one in the past few years.

EXPERIMENTS WITH AI AND DESIGN

On top of the design process research, I did several experiments with currently available AI tools to explore opportunities and limitations. This section highlights two key findings: prompts cannot capture the full complexity of ideas further on in the design process, and designers use different terminology than AI models, both aspects cause a lack of control for designers using image-based AI.

This section discusses one experiment for each of the seven tools mentioned: Midjourney, GPT-4, Vizcom, KREA, Luma Labs, Stable Diffusion - Dreambooth, and Stable Diffusion - ControlNet. For each experiment, I answer five questions: why the tool is interesting, what the task was during the experiment, who was involved, what opportunities and limitations arose, and how the tool could be used in practice.

For more output from the experiments, check appendices A-H

Figure 1.3. Three images generated using the tool Midjourney.

MIDJOURNEY

Perhaps the most popular image-generating AI tool, Midjourney, made the news when an AI-generated image won an international digital art competition in 2022 (Roose, 2022; Midjourney, n.d.). Designers at Gazelle were interested because Midjourney is considered to be the best text-to-image generator currently available for consistent high-quality results (Guinness, 2024). They had previously experimented with Midjourney and found several interesting results but are unsure of how to use it effectively considering their needs.

I asked one designer at Gazelle to design an electric urban fat bike using Midjourney based on a simplified fake proposition describing a market trend, target group, and design goal. The goal was not to come up with a finalised design, but purely to see how the designer interacts with the tool during the explorative stage of a design process.

During the experiment, the designer proceeded with a trial and error method, some generated images were great, others disappointed. Nevertheless, Midjouney is very helpful in the beginning stages when the designer explores multiple directions. He marked images that he liked and neglected the ones he disliked. At some point, he tried re-iterating the preferred images and using reference images he looked up on Google. This helped him communicate the direction he aims to go. Midjourney followed, however, at some point, the designer wanted to change specific details or shapes and failed to make MidJourney understand the shapes he proposed. This is the moment the designer lost interest and said he would normally switch to sketching or possibly Vizcom (Vizcom, n.d.).

When working with Midjourney, the designer mentioned that his perception of terminology seems to be different from how Midjourney interprets it.

Additionally, he mentioned he found it difficult to create inspiring images without steering away from archetypes too much. For example, at first, the designer used "fat bike" which led to fat bikes with very generic frames, later he dissected "fat bike" into "bicycle with fat tyres" or, when using a guiding image, into "two-wheeled vehicle".

The designer described the process as useful for quick inspiration, but not necessarily an improvement to the ideation or inspiration without AI. It does not work too well when iterating further on images because he cannot control what will be kept and what will be changed.

Figure 1.4. Three images generated using GPT-4.

GPT-4

OpenAI has released the massively popular ChatGPT and DALL-E. Since 2023 these have been combined in GPT-4 (Achiam et al., 2023; GPT-4, n.d.). GPT-4 allows users to give textual input but also allows different modalities like images, data files, and PDF documents. It can also generate images and graphs as output. This allows for a multi-modal collaboration, unlike most text-to-image generators that only use the textual and imagery modality.

During two short experiments, one with and one without a designer, the goal was to generate e-bike designs fitting the style of Gazelle. Because I was unsure how to best explain the design style of Gazelle to GPT-4, I asked it to give me a definition. The reply contained several fitting terms but remained top-level using generic terms like "robustness", "comfort", and "integrated components". This translated to basic designs that did not necessarily fit Gazelle's design style.

The strength of GPT-4 lies in the combination of a large language model, which is tuned to understand the textual input from the user and directly translate this into improved prompts for the text-to-image functionality of

DALL-E 3. Additionally, GPT-4 can offer a second opinion on a design made if the user uploads images to the AI. Limitations are similar to Midjourney, there is a lack of control. Additionally, it does not have a strong reference image functionality. It can however be used to do research, offer simple feedback, and generate inspiration.

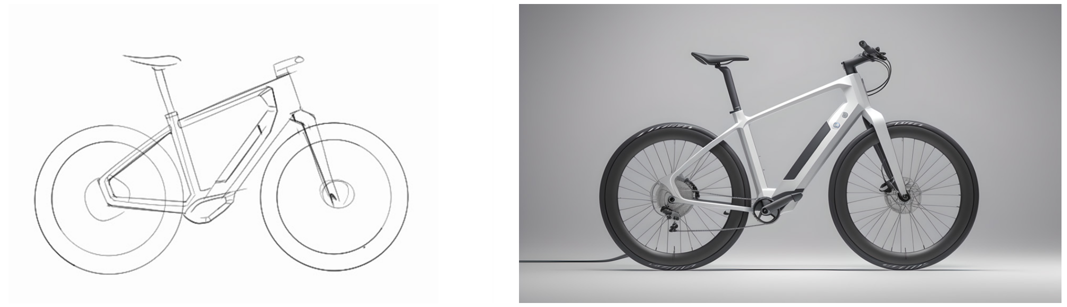


Figure 1.5. A sketch of a bicycle rendered using Vizcom.

VIZCOM

In July 2021, Jordan Taylor and Kelan Richards founded Vizcom, a sketch-to-render AI tool which aims to skip the tedious and time-consuming rendering and shading steps (Aquahug, 2021). Ultimately, they aim to go from sketch to 3D file so designers and artists can edit the 3D file in programs like Blender or Gravity Sketch.

Vizcom started in an automotive design community, therefore, the model is primarily trained on automotive images, clearly steering product renders in that direction (Automotive Design Planet, 2021). By now, the algorithm can do more than just automotive and is also being used for different product design categories and architecture.

During several experiments with a designer at Gazelle, he used line drawings with proper geometry as input and received rendered drawings from the program. He then tried to render these so the stakeholders could better understand design directions. The programme did not always properly understand his input and also did not always render it in the way that he wanted. Once again, the AI lacked control.

What Vizcom does significantly better than competitors like PromeAI (PromeAI, n.d.), OpenArt (OpenArt, n.d.), RenderAI (RenderAI, n.d.), and even remix mode in MidJourney (Midjourney, n.d.) is the workspace it offers. Vizcom allows users to easily sketch and resketch inside the program. This allows for making quick adjustments on the original and generated images to modify the output step by step-and-section by section. Apart from this strength, Vizcom does offer the possibility to regenerate only parts of the image, but the generative process still lacks proper textual control. Even with both text and image input, the AI cannot consistently generate output that fits the designer's envisioned concept. Additionally, using Vizcom requires too much time for quick inspiration in the very beginning.

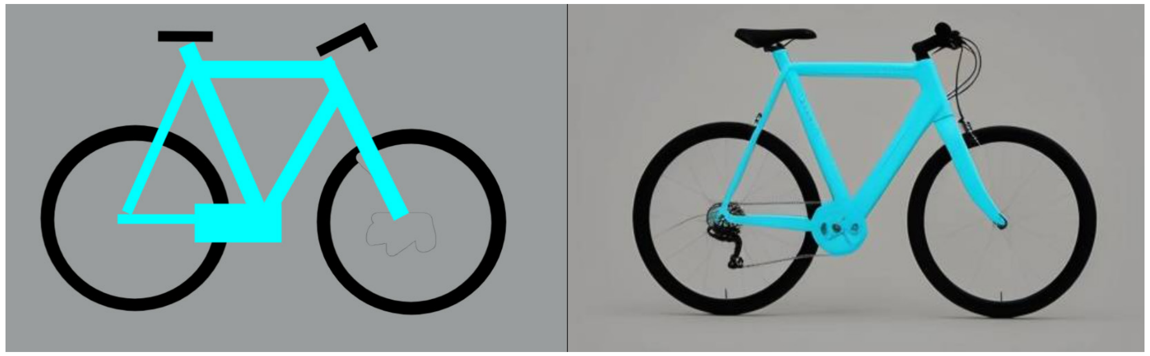


Figure 1.5. A very simple bicycle approximation in the live render interface of KREA.

KREA AI

In March 2022, Diego Rodríguez and Victor Perez founded Krea AI (Krea AI, 2023). This is an AI platform with multiple tools. It has AI patterning, logo illusion, and upscaling and enhancement of images. However, the most interesting feature is the real-time generation. Most image generation tools function based on a latent diffusion model, Krea AI, however, functions based on a latent consistency model. Latent consistency models speed up the iterative generative process by predicting the generated outcome directly in latent space instead of diffusion step-by-step.

The real-time generation tool allows users to render their visual and textual input in real time, meaning that changes on the canvas immediately influence the output window. This can be done through the control window on Krea.ai, but you can also use a camera or screen capture input, this enables users to use their preferred sketching platform like Photoshop or Procreate, and immediately render their changes. The tool allows users to set the AI's control. Setting the AI to be more influential could be useful for inspiration, but setting the visual input to be more influential allows the user to have more control over the visual details of the render.

During an experiment, I asked myself to create variations to the Avignon, one of Gazelle's e-bikes. I used a picture of the Avignon on top of which I drew using the app Concepts (TopHatch, n.d.) and streamed to Krea AI using Microsoft Teams (Microsoft Teams, n.d.). I got a lot of inspiration from the Krea output, however, I could not save it as it immediately changed if I made any minor changes. Nevertheless, the direct rendering tool is a great way of additional direct inspiration when sketching. The designer is still in control because one still has to sketch to explore the right shape. Krea does not interfere with the natural design process, it does complement it. Enabling users to look back at previously generated renders would be a useful feature, additionally, the latent consistency functionality strongly pulls the generated output towards training images. If the model is too strong, it is difficult to get a unique output.

Figure 1.6. 3D objects generated using the Genie tool by Luma Labs. The objects follw from the prompts "bicycle frame", "red electric bicycle", and "bicycle saddle".

LUMA LABS

The models above have focussed on text and images, but AI is also rapidly moving into the 3D realm. Luma Labs is a company that focuses on the 3D workspace, founded in 2021. Their first tool enables people to scan objects with a phone and turn these scans into 3D files (Luma AI, n.d.). Their second tool, called Genie, generates 3D models based on text prompts (Singh, 2023).

The scans made by Luma are both 3D and in colour. The finish and reflections seem to be taken into account when capturing an object. This offers great possibilities for CMF renderings. Imagine saving a product finish you encounter in 3D and applying it to another product directly from the app.

I tried to generate multiple bicycle and bicycle frame designs using Genie. Although I did generate objects recognisable as what the prompt described, the output was too generic to offer any relevant inspiration. The possibility to generate a model based on a text prompt skips many steps in the traditional design process. Generating a 3D model from text directly skips the explorative and exploitative balancing process which is fundamental for an interactive design process. The designer does not slowly get an understanding of what a final idea should be like, instead, he judges if fully rendered ideas are fitting or not, without adjusting or reframing the design goal between the start and end of the process. This is a clear example of the lack of control that comes with image-generating AI. One could argue that the 3D shapes offer inspiration, and they do, however, the 3D files are not of a resolution where they offer significantly more inspiration than 2D visualisations would.

Figure 1.7. Three images generated Stable Diffusion with Dreambooth trained on images of the Grenoble.

STABLE DIFFUSION - DREAMBOOTH

The third text-to-image generator that rapidly gained popularity is Stable Diffusion. It launched on August 25th, 2023 and is a text-to-image latent diffusion model created by CompVis, Stability AI, and LAION (Rombach et al., 2022). Stable Diffusion is unique because the code is publicly accessible; it also allows people to modify the code and train it based on different datasets. Alternative to retraining the full model, people train smaller models within Stable Diffusion. This is what DreamBooth does. It allows people to train the Stable Diffusion model on a separate dataset with a new label. This way, Stable diffusion can generate pictures related to a very specific and unique label (Ruiz et al., 2023). This way, the user has more control over what he wants to generate.

After multiple attempts to train Dreambooth on datasets with either the same or different model Gazelle bikes, I got mixed results. Some output images are very consistent and close to reality, others show deformed bikes or objects that are only vaguely recognisable as bikes. If the dataset is big enough or the product is not too complex, Dreambooth could be used to generate variations with a given style. Bicycles prove to be too complex for the model to comprehend. Stylistic features are taken into account, but the detailing is not sufficiently different for it to clearly offer inspiration. The user needs to critically look for inspirational features, they are not apparent and usually not present.

Figure 1.8. The base image, the inpainting region, and a generated image by Stable Diffusion with ControlNet: Inpainting.

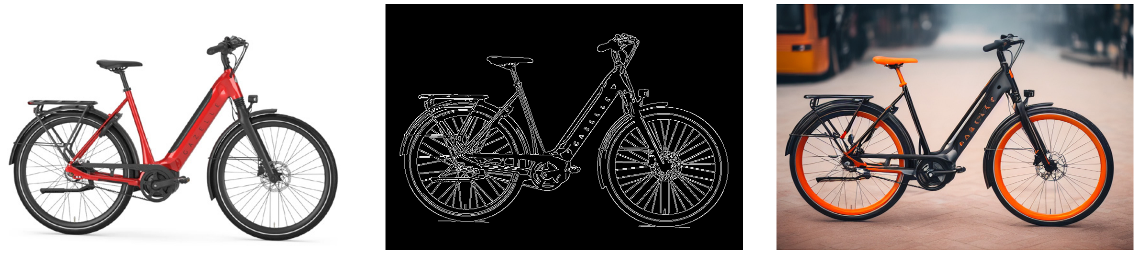


Figure 1.9. The base image, the canny image, and a generated image by Stable Diffusion with ControlNet: Canny Image.

STABLE DIFFUSION - CONTROLNET

ControlNet is another add-on for Stable Diffusion that offers the user more control. It is a model that allows the user to condition the Stable Diffusion model using an image. This enables greater control over image generation without the trial-and-error process of prompt engineering (Zhang et al., 2023).

The canny image feature uses a line drawing representation of hte original image to generate variations. While some lines in the canny image indicate outlines, others indicate material contrast. The image generator does not make a distincntion between the two. Additionally, the shading does not mimic Gazelle's form language, the tool does not offer sufficient control here.

When testing the inpainting feature, I aimed to generate only the frame while keeping the bike geometry constant. Inpainting allows the users to indicate a specific area for the model to generate. When I indicated the area for the frame, Stable Diffusion generated only a new frame, while the geometry and all other parts remained the same. The generations did mimic the original image with variations on the original frame. The results were high-quality, possibly because the complex parts of the bike, like the spokes, gears, and wiring, did not need to be generated.

This was one of my favourite AI tools because it allowed me to ask for variations in an indicated area. This offers me inspiration exactly where I need it. The UI was not ideal but could easily be improved. Especially when generating complex products, this enables Stable Diffusion to generate components that are comprehensible to the model.

CROOKED BOWTIE EFFECT

What all tools above have in common is that they use text as input and images as output modality. Some tools use both text and images as input and also have both modailties as output.

The saying goes: "A picture is worth a thousand words", generally attributed to Arthur Brisbane from the Syracuse Post in 1911. If a picture is worth a thousand words, how could we write a 20 or even 50-word prompt that explains the full complexity of what we want a generated picture to look like? That is exactly what the Crooked Bowtie effect explains (Verheijden & Funk, 2023). All text-to-image generator face this problem.

When the complexity of an idea exceeds the complexity that a prompt can capture, the generated image will not capture the same complexity either. This is what Figure 1.10 illustrates as what Verheijden and Funk (2023) call the "crooked bowtie effect".

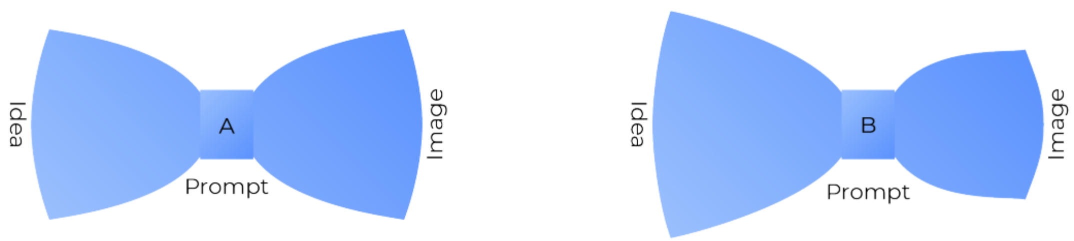


Figure 1.10. "Visual representation of crooked bow-tie effect. The image generation process can be seen as a bow tie. When used for inspiration (A), an idea converges into a prompt, from which an image is generated. Both sides of the bow tie roughly feature equal richness, just in different formats. When used for final designs (B), the bow tie becomes crooked on the left. Now, a very rich and complex concept must be converted into a prompt, from which it is very difficult to generate an equal amount of richness or accuracy." (Verheijden & Funk, 2023)

AI TOOLS THROUGHOUT THE DESIGN PROCESS

To visualise the AI tools opportunities, I made Figure 1.11. It illustrates how different tools fit in either the framing, envisioning, realising, or validating phase and if they are useful during explorative or detailing work. However, it is important to mention that there is no clear method or norm for using AI in the design process. These are the results from my experiments, where I think the tools might be useful, but people must keep experimenting with tools to explore new applications and use cases. Nevertheless, most tools fit in the upper section: the explorative phase. During the explorative phase, the designer learns to understand the problem and has fewer requirements. Towards the detailing phase, the designer needs more control to generate in a specific direction fulfilling several requirements. As mentioned before, most tools give insufficient control for these detailing steps.

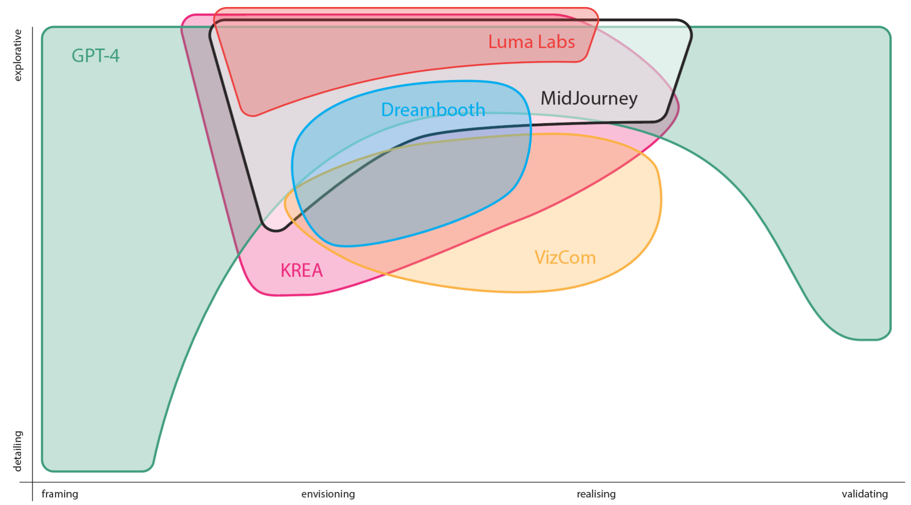


Figure 1.11. An indication of where several AI tools can easily be implemented. It shows the explorative versus detailing stage of a design project on the vertical axis, and the framing, envisioning, realising, and validating steps on the horizontal axis.

Apart from control, the intrusiveness of implementing AI in the design process can be a big obstacle. For example, KREA can be run on top of a current schedule, it does not interfere with the current process but does offer additional inspiration. Midjourney or GPT-4 however, are additions which the designer needs to find time for, they need to alter his usual process to use those tools. Therefore, designers might be more hesitant to use those tools. Additionally, Vizcom has an easy and intuitive user interface to quickly iterate using textual and visual control giving it an additional competitive edge. These considerations influence how likely designers are to adopt new tools.

MODELLING AI AND DESIGN IN PARALLEL

To better understand the similarities between AI tools used in design and the design process itself, I created a model that is an abstraction and simplification of the process. The models help us understand the tests and methods proposed later in this report.

None of the models take the context into account. The context is the space where the designer creates a design. That is an external environment which can randomly influence the designer. Including context in the models makes the models too complex to comprehend and will therefore be neglected. In most modelled tasks, I kept the context's influence to a minimum by using a constant surrounding and short time frame.

THE DESIGN PROCESS METAPHORICALLY

When the designer initiates the process, he aims to determine what he will design. He only knows the domain presented in the proposition. He gets an idea of what problem needs to be solved or what needs should be met. In the beginning phase, the designer exposes himself to as much inspiration as possible. Inspiration can come from various sources, including trend analysis, customer and market research, aesthetic trends, etc. Designers learn to find inspiration through these sources actively but also learn to be open to serendipitous inspiration. By exposing themselves to inspiration, they acquire the building blocks to form the foundation to start sketching. The sketching process is the beginning of direction within the given domain. Somewhere along the exploratory sketching process, the designer creates ideas of things that he deems as fitting to the problem. That is when the designer creates a mental model of what the final product should, and should not look like. They even form an unconscious understanding of what the product will look like, think of this as a blurry picture. During the rest of the design process, the designer uses multiple tools and methods to investigate his idea. He makes the image less blurry step-by-step. He first does this in 2D, later in 3D, and last in dimensionally bound 3D, until the image is fully in focus and defined.

MODELLING THE DESIGN PROCESS

All tasks can be simplified to the model below.

[tasks] * [mental process] * [input] → [output]

As mentioned, a complex proposition leads to a complex mental process. A complex task - the proposition in this case - or input - the research insights or mood board - requires a complex mental model to generate an output successfully. Complex mental models require large datasets to find the pattern. Datasets with all considerations to design work do not exist. Therefore, it is difficult to train an AI to predict the output of such complex propositions.

Nevertheless, designers and product managers would benefit from understanding designers' mental design processes to manage expectations. Modelling a simpler task could give insight into the mental model of designers and/or product managers. For example, if both parties would rate numerous images on different axes a model could learn their individual mental model containing how they rank images.

[rate image] * [mental model] * [images to be ranked] → [image rating]

The model would predict how different members rate images, which indicates their interpretation of both the image and terms on the axis. However, some images contain too many aspects to be consistently interpreted. In other words, if the "images to be ranked" are too complex, the mental model becomes overly complex as it needs to take into account too many different aspects. Alternatively, simple images with a very complex task also require a very complex mental model as the images might not have sufficient concrete features to be understood directly. This could cause people to interpret terms differently for each image.

The phenomenon above illustrates the ambiguity of mood boards and aesthetic visual-textual communication in general. If the pictures on a mood board contain multiple objects or features, they quickly become incoherent and ambiguous causing them to be freely interpretable. Possibly, AI could identify image fuzziness and make them more unambiguous and less sensitive to interpretation. That would enable designers to generate more coherent designs when presented with the same inspirational images or mood boards.

Quoted directly from the paper in this report:

"[image] labels describe low-level visual features, like colour and texture, and higher-level visual features, like style and emotion (Li et al., 2009). Especially the higher-level features are ambiguous since perception can differ between people or for a single person over time. Li et al. (2009) call this "fuzzy aesthetic semantics" which can also occur in the labels in AI training image datasets. This leads to ambiguous words in image labels."

This illustrates a key problem between both AI and its user, but also between humans themselves. Describing images and visual features is incredibly difficult because many people have different interpretations of both images and terminology. There is a semantic misalignment which makes it impossible for humans to intuitively understand AI. Just like humans miscommunicate, so does AI, but with AI it is more difficult to understand where the miscommunication originates from.

SHARED MENTAL MODEL

The goal is to use AI to enable unambiguous creative communication to facilitate the alignment of a brand's design style and product portfolio. To comprehend the problem, I created the visualisation as shown in figure 1.12.

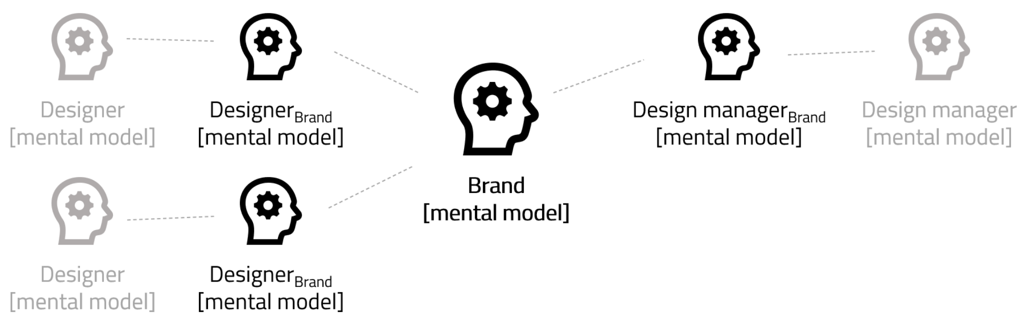


Figure 1.12. A schematic overview of the different mental models resulting from the brand mental model.

There is a brand mental model, which is the brand's visual style. Three people, two designers and one design manager, work with the brand mental model, however, they all have their own, differing, interpretations of the brand mental model. Additionally, this means they do not have the same perception of images and terms, which complicates the communication. The goal is to minimise the negative effects of all of these different interpretations.

PROBLEM STATEMENT

Combining the takeaways from both the design process and AI in design analysis leads to three key implementation opportunities:

1. Due to the safe development mindset within Gazelle, design innovation develops more through evolutionary than goal-oriented steps. Quick and broad generated output of AI could be used to explore directions and create low-budget vision concepts offering direction to Gazelle designers.
2. Even when combining textual and visual input, AI offers insufficient control to generate visualisations of rich and complex ideas. AI could be useful if it is tuned more towards Gazelle's design style and offers inspiration during the exploratory phase. A model could be tuned and implemented with a guiding method or roadmap to facilitate effective inspiration.
3. Misaligned perception and articulation of perception complicate the unification of the design style in the product portfolio. It also complicates concise communication with AI models. AI could facilitate unambiguous creative communication through reflection, unambiguous images, or group alignment to strengthen the vibe within a design team and enhance collective understanding of the Gazelle design style.

After considering these three directions side-by-side, option 3 came out on top. I reviewed the following 8 criteria named from most to least influential in my decision.

1. Useful for overall design quality - the added value towards a high-quality design according to Gazelle's design characteristics (comfortable, accessible, and quality) and the integration of design goals in a design.
2. Feasibility - the extent to which it is expected to be achievable within the timeframe, with the resources available, and with the skill sets involved during this project.
3. Useful for design efficiency - the extent to which it is expected to speed up the design process. Either through speeding up tasks or by making the process more linear.

4. Useful for PDP efficiency - the extent to which it helps to speed up the entire product development process through the integration of engineering considerations or by creating a more linear process.
5. Useful for brand recognition - the extent to which it helps Gazelle make their product more recognisable to customers. This can be through original outstanding designs or through creating a clear product family.
6. Relation to AI - the extent to which AI will play an essential role in this direction. A higher score resembles a bigger involvement of AI in the proposal because that would increase the benefits for the user.
7. Personal interest - the extent to which I am motivated and interested to pursue this direction.
8. Easy to code - the difficulty I expect in the coding part if I would need to code anything. Preferably, it is easy to code, or I can use off-the-shelf models to reach my goals because I did not have time to master coding during the project.

As mentionaed, option 3 focuses on strengthening the shared vibe. This option scored as the more prevalent and achievable project. There is a clear problem and a solution would widely benefit the PDP and design quality. It would also help the designers themselves. An important example of a misaligned vibe is in the use of mood boards, more on this can be found in Appendix I.

The final design problem is:

How could AI strengthen the vibe to align the perception and articulation of perception within design teams?

CONCLUSION

This chapter covers the design process at Gazelle and the opportunities and limitations of AI tools. Collectively, these lead to a direction of aligning perception and articulation of perception within design teams, paraphrased as strengthening the vibe. The direction follows from the designers' desire for a unified recognisable brand; the misaligned perception and articulation of perception among designers, the misalignment between designers' and AI's understanding of terminology; and the lack of control when using generative AI tools.

SOLUTION FINDING

The solution-finding section explores how to approach co-creative image labelling to effectively strengthen the vibe within a design team and between AI and designers. The experiment findings indicate that it does help align semantics within a participant group, but also sees the alignment fade 4 weeks after a labelling session. LoRA also proved to be a promising way to externalise a participant group's vibe, however, requires more images for effective training than the amount the participant group labelled. Currently, the model does not capture the full depth and complexity of the vibe.

First, the chapter explains an abstraction of the problem to make it more comprehensible. Second, it briefly explains three directions that could solve the problem described under problem-finding and why co-creative image labelling was chosen. The next section describes numerous considerations and preparatory research to set up an effective image labelling session. In the last section, this chapter goes into the findings from the experiment, mainly those the paper does not mention.

For information about the experiment specifically, read the paper "Co-creative LoRA model image labelling for semantic alignment in design teams".

KEY TAKEAWAYS SOLUTION FINDING

1. Image labelling works as a semantic alignment. The quantitative research indicates a noteworthy improvement in semantic alignment. Moreover qualitative research supports this statement, although the findings are not statistically significant due to a small participant group.
2. Collectively training a LoRA model does not capture sufficient complexity of a group's perception and articulation for it to be an effective externalisation. Groups must label a larger and more diverse image dataset before a LoRA can capture the full complexity of the group's vibe. This could be achieved but is very time-intensive.
3. Participants must keep using the semantics established within the image labelling session to remember the terminology. This helps them use the same terminology. Additionally, the perception of images quickly misaligns if participants do not stay engaged with the terminology and each other's perceptions.

THREE SOLUTION OPPORTUNITIES

There are three main solutions to the previously mentioned problem (the other considered options have been neglected in this report). First, designers could reflect on their own perception by training an AI to understand and illustrate their personal mental model. Secondly, designers could align the group's mental model by collectively training an AI which would force them to align. Last, designers could train an AI to understand when an image is ambiguous and use it to generate unambiguous communicative visuals.

REFLECT ON PERSONAL MENTAL MODEL

AI as a reflective partner has previously been suggested and proven to be useful (Van der Burg et al., 2023). AI can help designers become aware of their "own implicit perspectives". If designers learn to understand their implicit perspectives, they will have more knowledge to explain how they interpret things. Van der Burg et al. (2023) externalise individuals' implicit perspectives by training an AI to copy them. Looking into these externalisations could help team members better understand how their colleagues perceive different images. This way, AI could be a tool to increase mutual understanding between team members while maintaining individual perceptions and therefore the benefits of having a wide and diverse design team.

ALIGN MENTAL MODEL

Alternatively, by training an AI to understand individuals' perspectives, one could force a group to agree on the group's perspective and train an AI to understand it. Van der Burg et al. (2023) asked participants to individually label images, if a group labels images collectively, they must agree on what labels to use and what these labels entail. Low-Rank Adaptation (LoRA) models function on small datasets which can be labelled manually. They function within Stable Diffusion to generate images based on the LoRA model. A group could be tasked to individually label the images and validate if the labels are unanimous by generating images with the LoRA model. Collectively training the LoRA forces the group to align perceptions before externalising them in the form of a LoRA model. Externalising design interpretations and perceptions could be crucial in maintaining a company's intellectual property, even when key designers leave the company.

USE UNAMBIGUOUS IMAGES

The last two options focus on aligning and understanding the perception of imagery. However, AI could also make imagery less ambiguous. Some images can be interpreted in many different ways, while other images are more clear-cut and less ambiguous. For example, a designer might find an image that he likes because it communicates the right design style, but there are also some aspects of the image that he does not want to show. He could use an AI to generate a different version of that image and exclude confusing or ambiguous aspects like a split line, colour, materialisation, etc. Figure 2.1 previews a quick test with image variations to minimise ambiguity.

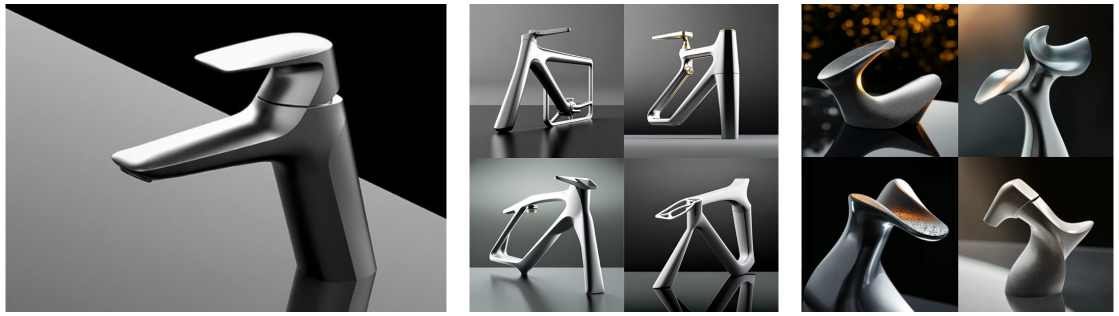


Figure 2.1. The left image shows a faucet design by Joris Wegner, the second image shows bike frames created by Stef de Groot using Midjourney with the faucet as a guiding image. I generated the last image with Midjourney and the faucet as guiding image, it shows abstract art based on the faucet.

CHOICE FOR LORA LABELLING EXPERIMENT

The LoRA image labelling session forces the designers to align their perceptions and articulations of perception before they can train a LoRA model they all agree with. This is the most direct option to enable unambiguous creative communication and facilitate alignment of a brand's design style and product portfolio. The other options could work but most likely would not be used in practice as they are too time-consuming and focus on helping individuals instead of the group, in which the process really lies. The image labelling session could be done once every few months and should semantically align a team for that period of time.

EXPERIMENT PREPARATION AND CONSIDERATIONS

As mentioned before, some images have a great amount of aspects that can be perceived. If the image has too many aspects, people will look at many different things and will therefore interpret an image differently, mainly because they are not really interpreting the same things. If an image is too simple and vague, people do not really know what to do with the image and will interpret it differently because there is too little concrete or relatable information.

The images in the labelling session should be informative to the extent that the designers can discuss the topics that are relevant to them within their day-to-day design work. Additionally, I want images that are not overly ambiguous. Therefore I talked to the design manager to set out which aspects must be included in an image to facilitate relevant discussions to semantically align the relevant topics.

IMAGE TYPE SELECTION

I looked into different categories that designers use on mood boards. I differentiated based on the complexity and number of differentiating aspects within an image. The categories can be seen per layer in Figure 2.2.

1. Non-functional form sculptures
2. Greyscale form studies
3. Form studies with colour
4. Product close-ups
5. Brandless products

In a questionnaire, I asked participants to rank each image on a scale from organic to geometric as this is a relevant feature for the designers. To check how consistently the images were interpreted, I took the standard deviation as an indication of the spread. A smaller standard deviation indicates a smaller spread and thus a more consistent interpretation. As portrayed in figure 2.4, the second category scores the most consistent. However, this category contained two pictures with very polarised scores, meaning an average score very far from the perfect middle score of 100. As Figure 2.2 indicates, a more polarised score correlates to a smaller standard deviation. To check, I took the average without standard deviations of pictures with an average polarisation bigger than 2 times the standard deviation in polarization. Therefore, it now excludes the outlier values based on extremely high polarisations, like those in the bottom right of Figure 2.3. Nevertheless, the second category still scores the lowest. The results are not statistically significant but indicate that the second category is the least likely to be ambiguous.

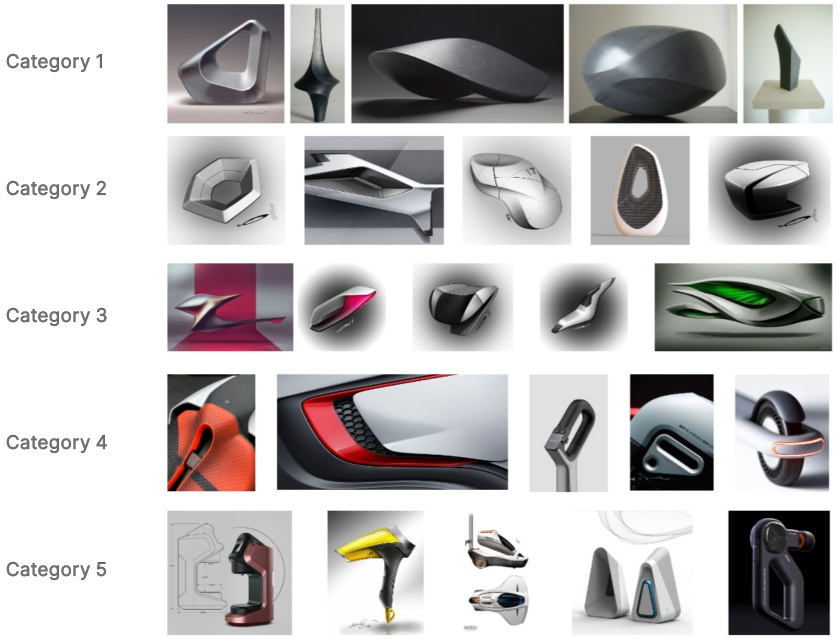


Figure 2.2. Each row above shows the pictures for the catagories as mentioned earlier. The top row shows non-functional abstract sculptures, the second shows greyscale form studies, etc.

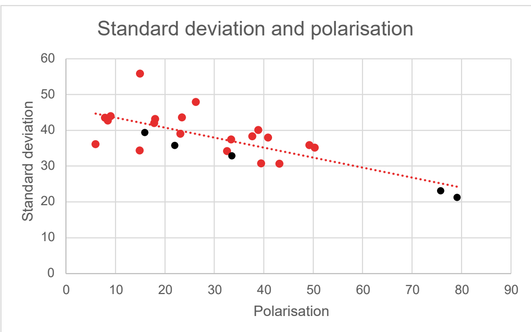


Figure 2.3. This scatterplot shows the polarisation on the horizontal, and the standard deviation on the vertical axis for 25 images. The black dots represent the images from the second category. This graph clearly shows the two outliers in the borrom right corner.

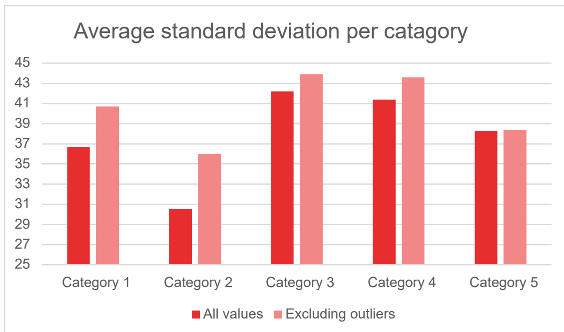
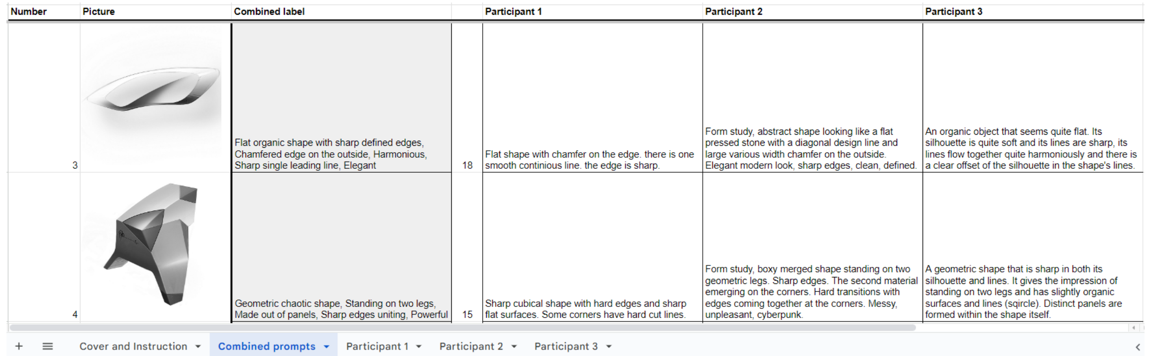


Figure 2.4. This clustered column graph shows the average standard deviation per category. Each right bar shows the average standard deviation without outliers with a polarisation bigger than twice the average polarisation.

SESSION FORMAT

During the session, participants will get a dataset with images of form studies. They have to write labels, these are textual descriptions, for each image. After writing individual labels, they will combine the individual labels into collective labels, one for each image. The goal of the session is not to label as many images as possible. That would be essential to make a proper externalisation of the team's knowledge, but the labelling session is primarily meant to align the aesthetic semantics within the team. Therefore, the designers must first formulate their individual perception in the image label, this functions as a backbone and trigger to include every participant's perception in the collective label.

The image dataset was separated into sets of 5 images, participants labelled individually first, and then combined with their labels for these images. This enables participants to thoughtfully formulate their own interpretations and find their flow. I considered letting participants write all labels before the session or at once during the session, but this would not allow the participants to adjust their labels throughout the combining process when they form their collective vocabulary.



Number	Picture	Combined label		Participant 1	Participant 2	Participant 3
3		Flat organic shape with sharp defined edges, Chamfered edge on the outside, Harmonious, Sharp single leading line, Elegant	18	Flat shape with chamfer on the edge. there is one smooth continious line. the edge is sharp.	Form study, abstract shape looking like a flat pressed stone with a diagonal design line and large various width chamfer on the outside. Elegant modern look, sharp edges, clean, defined.	An organic object that seems quite flat. Its silhouette is quite soft and its lines are sharp, its lines flow together quite harmoniously and there is a clear offset of the silhouette in the shape's lines.
4		Geometric chaotic shape, Standing on two legs, Made out of panels, Sharp edges uniting, Powerful	15	Sharp cubical shape with hard edges and sharp flat surfaces. Some corners have hard cut lines.	Form study, boxy merged shape standing on two geometric legs. Sharp edges. The second material emerging on the corners. Hard transitions with edges coming together at the corners. Messy, unpleasant, cyberpunk.	A geometric shape that is sharp in both its silhouette and lines. It gives the impression of standing on two legs and has slightly organic surfaces and lines (sqircle). Distinct panels are formed within the shape itself.

Cover and Instruction | Combined prompts | Participant 1 | Participant 2 | Participant 3

Figure 2.5. An overview of the excel sheet with the labels for the first two pictures.

The labelling happens in a Google Sheets file (Google Sheets, n.d.). The participants first fill out their individual labels in the participant-specific tab. The participant-specific tabs only show the images and a cell for one label per image. After each five images, the participants look in the "combined prompts" tab which shows the labels from all participants. They can look at each other's labels and combine them into a single label. Google Sheets allows users to fill out their personal labels in their individual tab without bias and facilitates the collective formulation of a group label for each image.

SESSION ANALYSIS METHODS

To analyse the effect of the labelling session, we gave the participants three questionnaires, one before, one after, and one four weeks after the labelling session. In these questionnaires, participants rank 25 images from fragile (0) to powerful (200). These questions will indicate the spread of perception. Each of these questionnaires also asks participants to describe two images textually, these descriptions will be compared to see if participants articulate images more similarly at different moments in time. The images are all form studies and participants never rank or describe the same form study twice.

During the session, participants work on scratch paper and in Google Sheets which can be analysed afterwards. Additionally, the group discussions get audio recorded for later analysis.

At the end of the session, the participants reflect using three key questions:

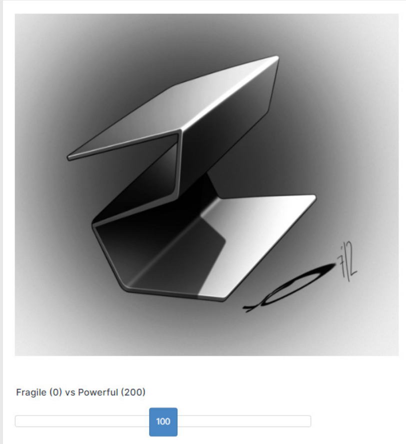


Figure 2.6. An example of the question where participans rate an image from fragile (0) to powerful (200).

1. How did you experience labelling the images?
2. How did you experience discussing the labels?
3. Do you feel that you have a better-aligned interpretation of design forms?

Additionally, all participants receive a post-session reflection questionnaire in which they share their labelling session experience using Likert scales (1-7).

EXPERIMENTING WITH FORM STUDY LORA

I experimented with LoRAs myself before facilitating a session on LoRA labelling with participants. I trained a total of four LoRA models using Hollowstrawberry's LoRA trainer (2023) in Google Collab to experiment with functionality and the effect of training methods. More output from the models can be seen in Appendix M.

In the first test, I used a dataset of 25 form studies, I wanted to check if the LoRA model could recognise the style and generate useful output regardless of the labels I gave to the 25 images I trained with The form study images were all created by Yo Kobayashi, George Yoo, Jeff Sihombing, and Evan Reese (Kobayashi, 2013; Kobayashi, 2017; Yoo, 2010; Sihombing, 2012; Reese, 2016). I did not label the images myself but used automatic BLIP captioning (Li et al., 2022). The test was successful, the images came out fitting the style of the training data but due to the automatic captioning, I did have control over what type of form study I wanted to generate.

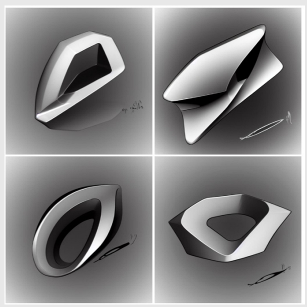


Figure 2.7. Output from the first test: a "form study"-trained LoRA model.

During the second and third tests, I wanted to increase the control when generating images, therefore I labelled the images using recurring terms to enable the LoRA model to analyse patterns. The goal was to see if I could generate images that fit the labels I gave in the training dataset. For example, I labelled 17 images as a "dynamic form" and wanted to see if the model could generate images that looked similar to the 17 images with a "dynamic form". The results were decent, but the size and diversity of the dataset really impacted the quality. During the first test, I used 23 images randomly picked from all the form studies I had downloaded. During the second test, I used 27 images I had selected so they would represent a diverse set of characteristics. The second dataset was slightly bigger and due to the diverse set I could make clear distinctions between different characteristics. The LoRA recognised some of them, but I did not have enough control to correctly generate images for all the labels I used during the training.

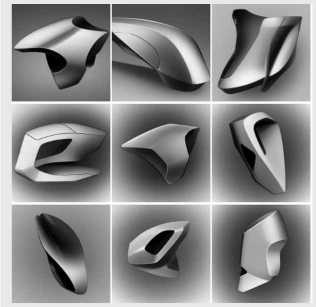


Figure 2.8. Output from the third test: a "form study"-trained LoRA model.

During the last test, I wanted to see if I could generate images of the front tube joint with different aesthetic values. I collected photographs of different front tube joint designs using Pinterest. This was a tricky process because there was not a lot of data available. In the end, the model underperformed probably due to the complexity of the pictures and the small size of the dataset (22 images). Some generated images were recognisable as front tube joints while others could barely be recognised as bicycle components. The tubes connected in odd and illogical ways. Additionally, the aesthetic values used in the training data like "matt frame" or "minimalist frame" were not translated to the generated images.

Figure 2.9. Output from the last test, a "front tube joint"-trained LoRA model.

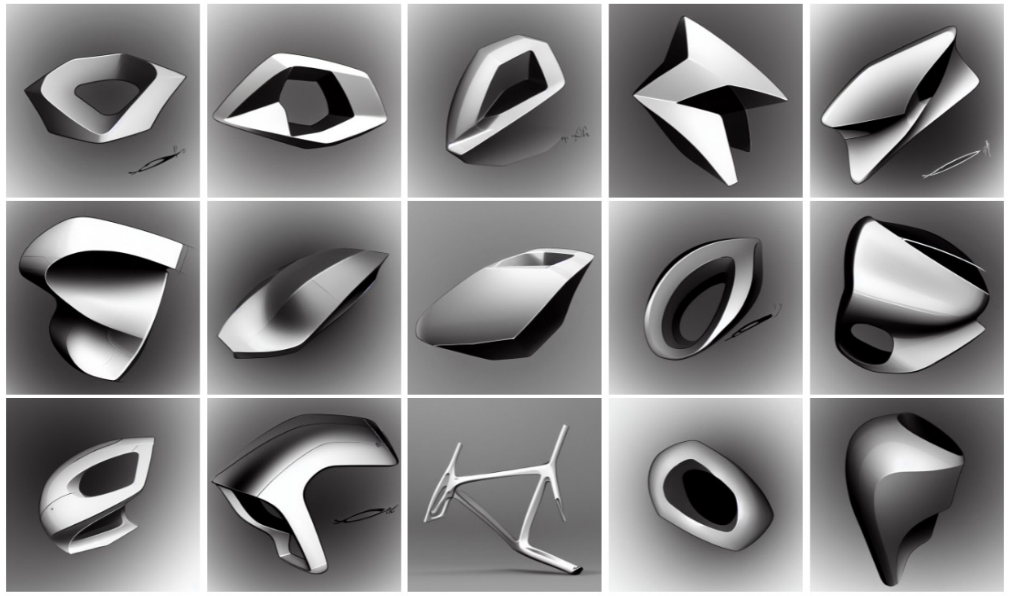


Figure 2.10. An overview of several images generated by a LoRA model trainedafer the third test. For the middle picture in the bottom row, the prompt specifically asked for a bicycle frame instead of a form study.

HOW TO EXPLAIN LORA LABELLING

Because the images and labels will be used for a LoRA model, the participants are curious to understand how LoRA models roughly work. I focus on three key things the participants should understand: that proper labelling helps the LoRA model understand patterns, using subjective terms is essential for the success of the labelling session, and that LoRA functions within Stable Diffusion which was also trained on a labelled dataset so that participants do not fight the base model.

Due to time constraints, I did not explain the full functionality of LoRA models. I clarified that the LoRA model works on top of an existing model and works best if labels occur in different images, then the LoRA can try to recognise the similarities and differences between images. Additionally, I explain that there is value in labelling subjective labels, both because models currently do not understand subjective terms, and because those are terms that are essential to align within the team.

To facilitate the explanation, I made a PowerPoint in which I gave a quick explanation of LoRA models and Stable Diffusion and gave several examples of image labels and LoRA-generated output. I deliberately chose not to use images similar to that of the database that the participants would label since that might influence how they labelled the images.

I did mention the terminology that I had discussed with the design manager as important terms that might be relevant to discuss for several images. Those helped steer the conversation throughout the image labelling process to relevant and insightful discussions.

LABELLING SESSION PILOT

Three master's design students joined a pilot session to check if the choices on methodology and session setup worked as planned. The session also highlighted where it could be improved. For the session, the explanation was given as planned and participants labelled the first five images. The image labelling process took a lot longer than expected. While the individual labelling process sped up as the participants looked at more images, the collective labelling process stayed slow as they struggled to combine all the descriptions.

After the session, the participants reflected on the labelling process but also on the session setup. There were a few important points of feedback:

- When considering handing the participants a cheatsheet with a few aspects they could think about, or split the description boxes into objective and subjective labels, they mentioned that the freedom to describe it helped them critically think about how they perceived the form studies.
- The participants were doubtful about the added value of the scratch paper. It might be an obstacle and obsolete. It was kept in place as it would offer extra data to analyse after the test. It offers insights if the participants also developed their perception individually.
- The participants said that showing five images per set was good, not too much or too little.
- The participants liked that there was no timer for when they were writing the labels, especially for the first few images because they needed to get used to the process of image labelling.
- The participants mentioned that if they looked at images for an extended period of time, they started noticing new aspects and interpretations.
- Lastly, the participants mentioned that it would be okay to steer the collective labelling process towards a final label if it carried on for too long. That is mainly because the test group noticed that they could talk about several images for a really long time.

SESSION RESULTS AND DISCUSSION

The full session results can be found in the paper. For readability reasons, the section below quotes parts from the paper. It also adds several findings and considerations from additional analyses and an additional questionnaire.

SESSION RESULTS

The session results can be seen in the results section of the paper. To quote the quantitative results:

> *"The participants rank 25 images before and 25 after the session on a scale from fragile (0) to powerful (200). The average scores before and after the session were 98.1 and 102.2 respectively. The rankings deviated from a perfect balanced score of 100 points by an average of 35.7 points pre-session and 37.0 post-session points. The average sample standard deviation, indicating the spread of the rankings, was 39.1 points pre-session and 27.3 points post-session. The average spread decreased by 30.1% when comparing the post-session results to the pre-session results."*

The results from the verbal reflection are:

> *"The participants reflect that mainly at the beginning they found it difficult to accurately and concisely label the images. After the first discussion round, they found the terminology to address the essential visual aspects of the pictures accurately. The group discussions were complex and long at the start, but the second group discussion was quicker as they aligned their terminology in the first discussion.*
>
> *The participants find the session interesting and insightful but it does become repetitive. Hearing other participants' perceptions of different images and why is especially interesting. The individual labels are the foundation for an insightful discussion. Because of time constraints, the participants labelled the 11th and 12th images as a group, without first writing personal labels. They reflect this is time-efficient but leaves less space to share personal misalignment between participants. They experience a contrast between wanting to label as many images as possible to create a better working LoRA model and wanting to take time to align their semantics."*

Last, the results from the reflection questionnaire with Likert scales indicate the following:

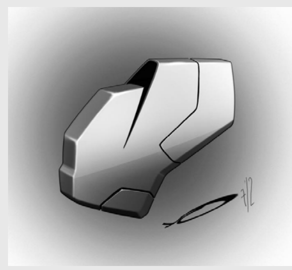
> *"The session helps participants better understand how they interpret form studies themselves, how others do, and how the group does. The participants also reflect that they started using a more and more similar vocabulary, nevertheless, they only indicated a slight increase in their ease in communicating how they interpret images. After the session, the group felt more aligned on the descriptions of form studies and felt that there was a better collective understanding of how to describe form studies."*

The descriptions given by the participants before the labelling session were quite different. Each description used a different format and varying terminology, and each mentioned some similarities but also many different aspects of the images. Figure 2.11 shows a form study with the descriptions from the pre-session questionnaire.

1. A rugby-shaped object with fine lines. It has two indents, with sharp cutouts
2. A pill-shaped object with the ends pointy like a rugby ball. There are organic split lines. A sharply defined hole which gives the impression of a handle and on the other side a gap with a floating part in it. Abstract shape, satin metal-like finish, strong, busy
3. This is quite an organic shape. Shapes are integrated and its lines really flow in a very natural way.

Figure 2.11. A picture with three descriptions written by participants in the pre-session questionnaire.

In the post-session questionnaire, participants described the image with more similar terms, focused on similar features, and were more aligned in which terms they gave to each form study. They followed a format which had been formed during the labelling session. On the other hand, they no longer described the overarching shape of the form study; words like "pill-shaped" or "rugby-shaped" were lost.



1. Geometric shape and silhouette with straight split lines. Harmonious. Powerful.
2. An angular shape. Geometric, with a relatively flat front. Flat surfaces. Chamfer on the botom. Straight angular split lines. Hard straight edges. Powerful shape.
3. A geometric shape that consists of quite an angular silhouette and sharp edges. Angular split lines flow over its flat surfaces. This shape is quite chaotic and powerful.

Figure 2.12. A picture with three descriptions written by participants in the post-session questionnaire.

During the session, participants realised why they used specific terms because they were forced to explain and support their labels towards other participants. For example, “I don’t think it is harmonious because this shape does not have the same parallel lines or rhythmic elements”. They also used field-specific terminologies and analogies, like saying a shape has a “Fibonacci-like” structure. The argumentation of terminology encourages participants to understand why they perceive and articulate imagery in a specific way.

LONG TERM QUESTIONNAIRE RESULTS

4 weeks after the labelling session, the participants filled out another questionnaire where they ranked 25 images and described 2 images. The long-term (4 weeks) average standard deviation was 43.7 compared to 39.1 and 27.3 for the pre- and post-session results. Oddly enough, there was a slight increase in the standard deviation, it even got consistently bigger.

The image descriptions did remain somewhat consistent. Two out of three participants remembered most of the vocabulary established during the session, however, used different words for the same shape. For example, participant 2 described the first shape as “geometric”, while participant 3 said it was “semi-organic”. In the second picture, participant 2 used “semi-geometric” while participant 1 called it “organic”. Participant 1 had forgotten most of the established vocabulary, and instead of describing the shapes in their own words, they shortened their description.

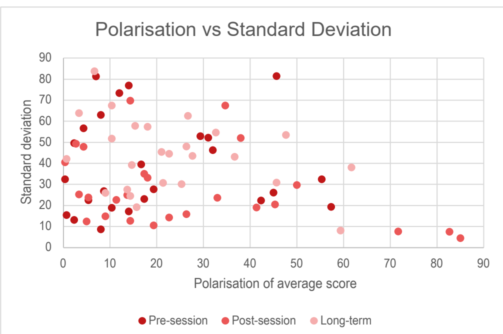


Figure 2.13. This scatterplot shows the polarisation on the horizontal, and the standard deviation on the vertical axis for 75 images. 25 images from each questionnaire: pre-session, post-session, and long-term questionnaire.

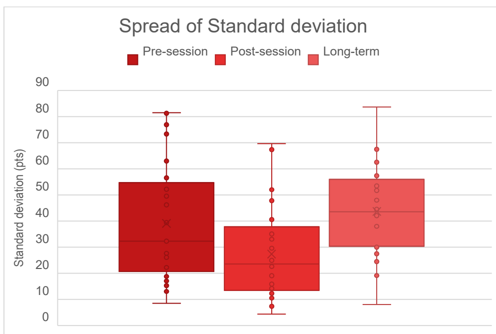
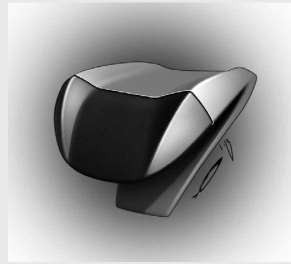


Figure 2.14. This box and whisker plot illustrates the spread of standard deviations per questionnaire. The average standard deviation in the post-session questionnaire is notably lower, then in the pre-session. The average standard deviation is highest in the long-term questionnaire.



1. Organic shape, chamfered edges, dividing split line, powerful
2. Semi-geometric shape, flat surfaces, sharp curved continuous edges, straight angular split line, folded shape, symmetric
3. A semi-geometrical shape with soft lines, flat surfaces, symmetric, split lines and contrasting panels. It is an elegant shape.

Figure 2.15. A picture with three descriptions written by participants in the long-term questionnaire.

POST-SESSION LORA MODEL

Using the data from the test, I did several tests to see if the LoRA could be seen as an externalisation of the team’s knowledge. I would only consider it a true externalisation of knowledge if the LoRA could generate images that fit the prompts. The biggest issue was that the design team only labelled 12 images during the session. This was sufficient to align their semantics, but probably not to create an effective LoRA.

During the first test, I tried training a LoRA as I had done before, but the results did not make a lot of sense. The shading was off, and some generated images that simply could not be imagined in 3D. The images were inconsistent, the LoRA seemed not to have sufficient data to sufficiently recognise the basic patterns, let alone the patterns linked to labels.

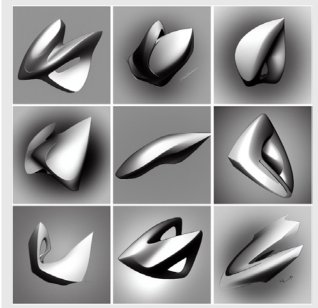


Figure 2.16. Stable Diffusion output using the LoRA model trained on the labelling session’s dataset

To make the LoRA perform more consistently, I took 213 unlabeled form studies by Kobayashi, Yoo, Sihombing, and Reese and put them in a regularisation folder (Kobayashi, 2013; Kobayashi, 2017; Yoo, 2010; Sihombing, 2012; Reese, 2016). LoRA Kohya trainer allows users to include a regularisation folder which helps the model understand the general image style and rules (Linaqruf, 2023). This made the images a lot more consistent, but it did reveal another problem with small training datasets, as shown in figure 2.17.

Possibly due to the limited amount of images with certain labels, generated images became quite polarised. If I wrote a prompt for a geometric shape, it generated extremely geometric shapes, the same for organic. The output does illustrate the key terminology correctly but does not capture the full nuance of the different terms, it could however be used as an indication.

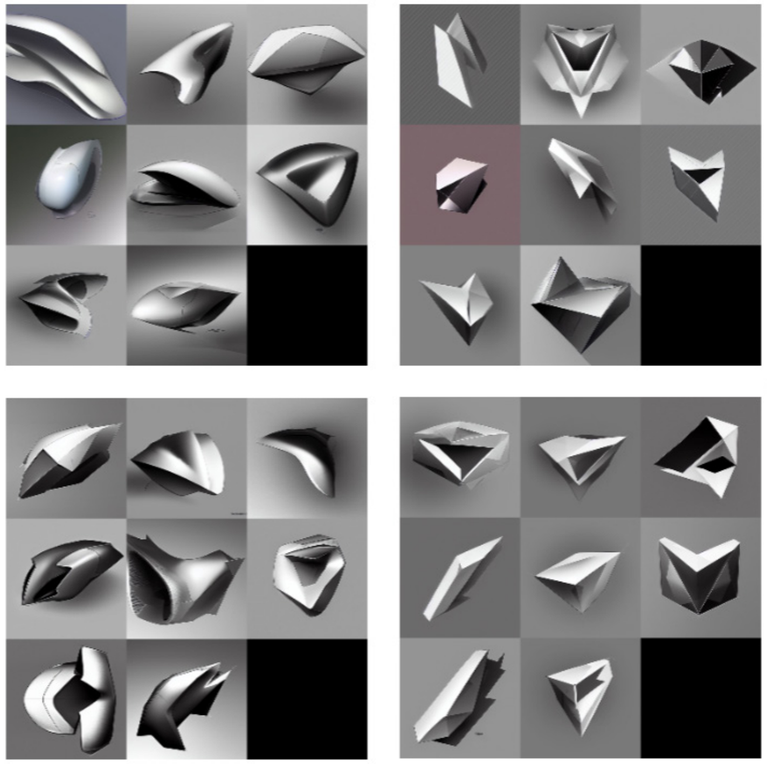


Figure 2.17. Stable Diffusion output using the LoRA model trained on the labelling session's dataset. The generated images strongly embody the key terms. There is no semi-organic or semi-powerful shape. The output appears to be very extreme.

CONCLUSION

This chapter goes into the development of the image labelling session. Collectively labelling an image dataset for a LoRA model forces designers to align their vibes before writing labels. Form studies were chosen as the most relevant image type to facilitate useful and on-topic discussions. The labelling session increased alignment in the participant group by 30.8% when comparing after labelling 12 images. The participants used more of the same vocabulary after the session to articulate their perceptions. After four weeks, the participant group lost the alignment generated during the session falling back to pre-session scores for perception. The post-session LoRA model generates output that generally fits the prompts but represents the extremes more than detailed nuances.

CO-CREATIVE LORA MODEL IMAGE LABELLING FOR SEMANTIC ALIGNMENT IN DESIGN TEAMS

Bouwe Theijse
TU Delft, The Netherlands
b.h.p.theijse@student.tudelft.nl

Vera van der Burg
TU Delft, The Netherlands
v.vanderburg@tudelft.nl

Derek Lomas
TU Delft, The Netherlands
j.d.lomas@tudelft.nl

This study explores the impact of co-creative image labelling on semantic alignment within industrial design teams, using low-rank adaptation (LoRA) models. The research was conducted with a three-person design team from a Dutch mobility company. It examines how co-creative image labelling could align semantics, thus perception and articulation of perception, among designers. Team members aligned their understanding and descriptions of images by engaging in co-creative labelling sessions. Alignment is critical for consistent and effective communication in design processes. After the session, participants interpreted images more similarly and used more similar terms to describe images. The study stipulates the potential of image labelling sessions as a tool to align semantics within design teams and transfer this alignment to AI image-generating models.

This study has been set up after a four-month collaboration with a three-person design team of a Dutch mobility company. The design team was curious about how they could use AI in their design process. During the collaboration, the designers tested numerous AI tools and evaluated how these influenced their design process. The study focuses on generative image-based AI tools, these are AI tools that generate images based on textual, visual, or other modality input.

INTRODUCTION

The rise of AI has triggered curiosity among designers about how they can use it. Especially since the introduction of ChatGPT, Dall-E, and Midjourney, generative AI has gained a lot of publicity. It is frequently said that AI will hugely impact the creative industries, one of which is the design industry (Tholander & Jonsson, 2023; Van Der Maden et al., 2023; Lawton et al., 2023). Van der Maden (2023) points out that AI could help designers "Generate, explore, and extend ideas faster by offering serendipity, surprise, and by generating friction through its unpredictable outcomes." Additionally, media illustrate the speed at which AI turn a text-based prompt into high-quality visuals (Roose, 2022; Davenport, 2022; Booth et al., 2024). On the other hand, Lawton et al. (2023) point out that this workflow is contradictory to the traditional understanding of how creativity functions: "Psychology tells us that, in creative processes, it is only through searching that we can discover what we were looking for all along" (Dorst, 2006; Goldschmidt, 1991). Given AI's opportunities and limitations, the question of how to use AI for design remains unanswered. This study goes in one specific direction that stems from the tests done by the design team at the mobility company.

BACKGROUND AND RELATED WORK

When testing the image-based AI tools, the biggest obstacle the designers faced was a lack of control when generating images. During the explorative steps, the designer slowly gets a better sense of the problem and the desired solution (Dorst & Cross, 2001). Therefore, the beginning of a design process is open, this is when the AI tools flourish since the designers welcome most input. As the explorative process develops towards a detailing phase, the designer considers a more refined solution space in which he evaluates the properties of the problem space, and vice versa (Dorst, 2006). The problem and solution spaces tighten due to a better understanding of each setting up numerous constraints. Later in the design process, designers need more control when creating. Their ideas become more complex, and so do their visualisations. However, when working with image-generating AI, the controlled translation of the idea toward a prompt for an image proves to have limitations.

Verheijden and Funk (2023) build upon this bottleneck: the prompt. A good prompt is the key to generating desirable images. Their research illustrates that an idea a designer imagines can be paraphrased to a prompt to generate a visualisation. When the idea is vague, most visualisations inspire the designer. However, when the idea is complex, the prompt never captures all essential details, this results in visualisations that only partially meet the idea's complexity. This fits the narrowing problem-solution space portrayed by Dorst (2006), AI tools do not offer sufficient control to generate images within that space.

Apart from limited words captured in a prompt, the lack of control stems from ambiguous image labels in AI training databases. Occasionally designers and AI models understand words differently. For example, the label "bike" is culture-dependent: some cultures might interpret it as a motorbike, while others might think of a bicycle. This problem has occurred before. Image labels have previously been used for search engines and art image retrieval. Labels describe low-level visual features, like colour and texture, and higher-level visual features, like style and emotion (Li et al., 2009). Especially the higher-level features are ambiguous since perception differs between people and for a single person over time. Li et al. (2009) call this "fuzzy aesthetic semantics" which can also occur in the labels in AI training image datasets. This leads to ambiguous words in image labels.

Words within a prompt can also have ambiguous meanings, they leave room for imagination and interpretation, what Xu (2021) calls "descriptive ambiguity". Designers can embrace descriptive ambiguity as serendipitous inspiration, however, it can also lead to frustrations when one aims to visualise an idea that is too complex to be captured by a prompt (Verheijden and Funk, 2023). Descriptive ambiguity and fuzzy aesthetic semantics are examples of semantic misalignment in which people articulate their perceptions using different words. Ideally, users and AI models are naturally semantically aligned and thus use the same terminology to articulate perception. This is one of the reasons OpenAI decided to relabel the training dataset for DALL-E 3 to produce more accurate results based on less ambiguous terminology (Betker et al., 2023). Relabelling the training dataset increased the mutual understanding of terminology between DALL-E 3 and the user.

The research around descriptive ambiguity and fuzzy aesthetic semantics (Xu, 2021; Li et al., 2009) mainly illustrates the semantic misalignment between users and AI. Humans themselves, however, interpret and articulate imagery differently, meaning they are not semantically aligned either (Biederman, 1987; Fei-Fei et al., 2007; Gilchrist, 2012). Hence, image labelling systems can never give captions that accurately fit everyone's perception (Karpathy & Fei-Fei, 2015). Some seemingly unambiguous terms will still be perceived differently by people from different cultures, professions, or generations (Wijaya & Yeniterzi, 2011). Image-generating AI models like Stable Diffusion or DALL-E are trained on massive datasets for a worldwide user base, hence they can never account for all perceptions of words. Two solutions follow: either the user learns to understand the model's semantics or the model learns to understand the user's semantics, both would create semantic alignment.

These findings identify two points of semantic misalignment: between designers themselves and between humans and AI. Each has its importance. First, semantic alignment among designers is essential when creating a strong and recognisable brand as proper communication forms the basis for design reasoning (Ranscombe et al., 2012). Additionally, designing a recognisable product family is essential for strategic brand management (Warell, 2006; Kreuzbauer & Malter, 2005; Park et al., 1991). Second, semantic alignment is the key to effectively communicating with, and using AI models. This research focuses on the first problem, aligning semantics within a design team.

Van der Burg et al. (2023) illustrate how labelling a dataset is an insightful process and how reviewing a self-trained image recognition AI illustrates the labelling party's perception of images. Design teams consist of multiple designers with individual perceptions of images and terms that differ between team members. Collectively labelling a dataset to train an AI model will force designers to agree on how they perceive images and articulate their perceptions unanimously. The process will semantically align the designers and the design team with the AI model they train.

Low-rank adaptation (LoRA) models allow users to manually label a small dataset and generate output within a bigger model (Gu et al., 2024; Hu et al., 2021; Li et al., 2024). Co-creative image labelling is when several people collectively label an image dataset. They must agree on how they use terms and what they describe for each image. If they train a LoRA model on the collectively labelled dataset, the model becomes an externalised version of their team's perception and articulation of perception. The image labelling session facilitates the semantic alignment within the design team.

PREPARATORY CONSIDERATIONS

Co-creative image labelling, when a group collectively writes labels for multiple images, should facilitate semantic alignment within the participant group to communicate how they perceive images. Before the participant group can label an image, participants must combine their individual perceptions and formulate what they perceive collectively to align on the words they use. This should increase their mutual understanding of perception and enable them to articulate it using the same vocabulary.

To facilitate an effective image labelling session, the images should not be overly minimalistic or complex. If an image is too vague, the image is likely to be ambiguous and no concrete labels can be given. If an image is too complex, too many elements can be seen for the participants to capture in a prompt.

Because all participants work as product designers and design 3D shapes, they will label a dataset of form studies. Normally, designers create form studies to explore shapes without focusing on functionality or colour. The selected form studies consist of a single shape, facilitating a focused discussion. To encourage all participants to share their perception of forms, they first label images individually before formulating the collective image labels.

STUDY SETUP

This section describes the research goal, method, questionnaire to measure results, and the image pre-selection.

RESEARCH GOAL

This research answers whether co-creative image labelling positively affects semantic alignment within the design teams.

The positive effect indicates that the semantics within the design team are more aligned, i.e, less varied. A questionnaire will quantitatively determine semantic alignment before and after the session. During and post-session observations will be evaluated qualitatively. The participant group is too small to claim scientific significance, therefore, the quantitative analysis will be used as an indication more than concrete proof.

METHOD

Before the starting the labelling process, the participant group gets an explanation of what LoRA models are, how they can be used, and how they can be trained. The explanation also illustrates how labels for two example images can have both objective and subjective tags. The facilitator mentions that the The participants are asked to label the images to the best of their abilities and are explicitly asked to focus on the subjective terms, like "fragile" and "powerful", because those require alignment most.

During the session, participants first individually label five images. Then they discuss each other's labels and formulate a single label per image. They repeat this for up to 25 images, or until the time runs out. Figure 3.1 shows the labelling process for a single image. A label includes a description of what can be seen in the image and how the participant or the participant group interprets the images. All participants are in the same room during the image labelling session. The facilitator shows the images on a screen and participants write labels on a personal laptop.

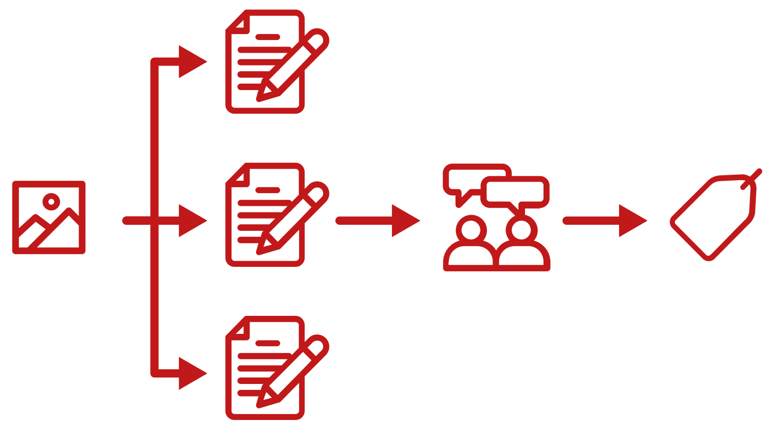


Figure 3.1. A schematic overview of the labelling process. First all participants see the image and write an individual description. Then they discuss their labels and formulate a single collective label for the image.

The participants reflect on the session as a group. They answer three key questions both individually and as a group.

1. How did you experience labelling the images?
2. How did you experience discussing the labels?
3. Do you feel that you have a better-aligned interpretation of design forms?

For the quantitative analysis, the experiment uses an online pre-session and a post-session questionnaire to measure the semantic alignment within the participant group before and after the image labelling session. Additionally, the participants' input gets recorded in digital spreadsheets and on pieces of paper, and the audio of their discussions and reflection gets recorded for qualitative analysis.

During the session, the facilitator can ask guiding questions to facilitate indepth discussion between participants.

QUESTIONNAIRE

To measure the effect of the labelling session, participants individually fill out an online questionnaire within 24 hours before, and within 24 hours after the labelling session. The questionnaire asks the participant to rank 25 images on a scale from fragile (0) to powerful (200). Additionally, the participants are asked to write two image descriptions. The form study images were all created by Yo Kobayashi, George Yoo, Jeff Sihombing, and Evan Reese (Kobayashi, 2013; Kobayashi, 2017; Yoo, 2010; Sihombing, 2012; Reese, 2016).

IMAGE PRE-SELECTION

The participants label up to 25 images that communicate semantics that the participants want to align on. In this study, the participants labelled form studies created by George Yoo, Jeff Sihombing, and Yo Kobayashi. Preparatory validation indicated that these images balance ambiguity and specificity resulting in a lesser ambiguous interpretation, and enabling participants to discuss limited but specific and characteristic aspects per image.

FINDINGS

The labelling session including an explanation and a brief reflection took 2.5 hours. The individual labelling took roughly 3 minutes per image. Writing the collective labels took about 7 minutes per image. Throughout the session, both the individual and collective labelling sped up drastically. The participants labelled a total of 12 images. The individual labels consisted of 21.3 words on average, and the collective labels consisted of 14.9 words on average.

SESSION OBSERVATIONS

As the participants individually label the first five images, they visibly struggle, occasionally adding or crossing things on scratch paper. The first few images also take notably longer to label than the last few images where participants were usually done before the advised two minutes per image. Between the participants, there was also a clear difference in working pace, similar to the length of image labels. The labels given in the Excel differ from the labels on scratch paper.

During the first discussion, participants realised how their labels differ and understood the importance of aligning definitions and terminology before writing the collective labels. For example, they ask each other how they should use “shape” or “surface” and whether the line is “sharp” or “flowing”. They learn that they use different terms for similar aspects in the image but also the same term for dissimilar aspects in the image. Especially the discussion on vague terms like fragile and powerful offers new insights between participants on why some labels are different. The participants consider subjective terms like “cubic” and “geometric” overlapping. The participants choose to use only “geometric”.

Language barriers also affect the labelling process. None of the participants were native English speakers but they labelled in English to connect the LoRA to the base mode. During the session, some participants asked for translations of words. Additionally, they were trying to find the right nuanced words. For example, they doubted words like “elegant” and “fragile” as the opposite of powerful.

One form of study is very similar to the design language of the mobility brand. All participants agree that it fits their brand and they all become more focused on finding the right words to caption the image. This label is slightly longer than average (18 words). They add a unique label as this form fits their brand so well, almost like saying “This is -brand-”.

The participants also make several verbalised realisations. For example, after discussing image labels for the second set of images a participant realises that they are “slowly getting the same vocabulary”. They also reflect on previous images and labels, realising some of the chosen terms still contain ambiguity: “It is powerful, but not in the way we labelled before”. To ensure they kept using the same terminology, one participant kept track of the most important terms and their antonyms as shown in Figure 3.2

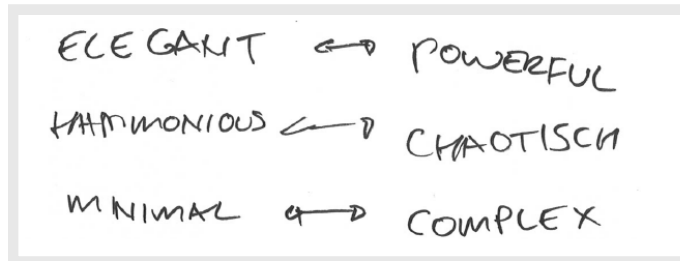


Figure 3.2. Jotted down notes indicating the opposites and focus points in the labels.

QUESTIONNAIRE

The participants rank 25 images before and 25 after the session on a scale from fragile (0) to powerful (200). The average scores before and after the session were 98.1 and 102.2 respectively. The rankings deviated from a perfect balanced score of 100 points by an average of 35.7 points pre-session and 37.0 post-session points. The average sample standard deviation, indicating the spread of the rankings, was 39.1 points pre-session and 27.3 points post-session. As illustrated in Figure 3.3, the average standard deviation decreased by 30.1% when comparing the post-session results to the pre-session results.

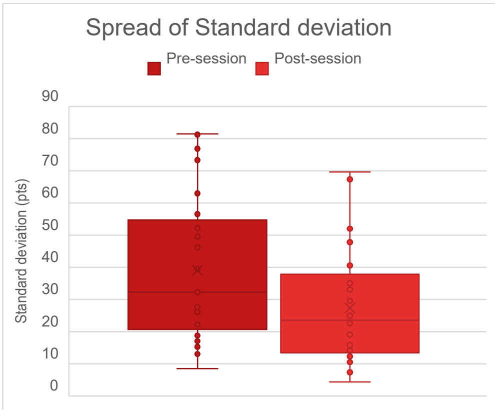


Figure 3.3. This box and whisker plot illustrates the spread of standard deviations for the pre- and post-session questionnaire.

The image descriptions given in the post-session compared to the pre-session questionnaire also indicate increased alignment amongst participants. In the pre-session responses, participants use a wide variety of words and do not seem to look at the same features of each shape.

REFLECTION

The participants reflect that mainly at the beginning they found it difficult to accurately and concisely label the images. After the first discussion round, they found the terminology to address the essential visual aspects of the pictures accurately. The group discussions were complex and long at the start, but the second group discussion was quicker as they aligned their terminology in the first discussion.

The participants find the session interesting and insightful but it does become repetitive. Hearing other participants' perceptions of different images and why is especially interesting. The individual labels are the foundation for insightful discussion. Because of time constraints, the participants label the 11th and 12th images as a group, without first writing personal labels. They reflect this is time-efficient but leaves less space to share personal misalignment between participants. They experience a contrast between wanting to label as many images as possible to create a better working LoRA model and wanting to take time to align their semantics.

The session helps participants better understand how they interpret form studies themselves, how others do, and how the group does. The participants also reflect that they started using a more and more similar vocabulary, nevertheless, they only indicated a slight increase in their ease in communicating how they interpret images. After the session, the group felt more aligned on the descriptions of form studies and felt that there was a better collective understanding of how to describe form studies.

DISCUSSION

The findings indicate three key things:

1. Participating in co-creative image labelling sessions benefits the semantic alignment in design teams. The spread in the post-session questionnaire's results was notably lower than in the pre-session questionnaire.
2. Throughout the session, participants labelled images more quickly. They used more of the same terminology suggesting that they find it easier to articulate their perception of images.
3. When reflecting on the session, participants mentioned that they were less likely to share their honest interpretations if they did not write them down first. This suggests relations within a participant group influence people's willingness to share their full perception. The individually written labels encourage participants to share their own and check other's labels thoroughly.

These findings motivate design teams to align semantics to facilitate effective communication and align aesthetics within their product portfolio. Improved aesthetic alignment leads to a more recognisable product family and indirectly to better customer binding (Kreuzbauer & Malter, 2005).

The idea of labelling a dataset to align personal and group semantics is similar to the educative functionality of explaining things to other people. It is widely accepted that explaining increases one's understanding of the material because one must first understand what must be explained (Annis, 1983; Coleman et al., 1997; Fiorella & Mayer, 2013; Duran, 2017). Kobayashi (2019) also suggests that explaining face-to-face is more beneficial in learning than explaining to oneself. Additionally, learning through discussions has a positive relation to performance, preferably in smaller groups as it increases participation (Ellis et al, 2004; Pollock et al., 2011).

To facilitate an on topic and relevant discussion, images in the dataset should be relatable for the participant. With the current dataset, the labelling could have been more focused on small nuances if all the form studies were similar to the company's design style because the participants would value a more accurate label. On the other hand, even though it is not the direct scope of the experiment, the labels might be misunderstood by the LoRA model because they focus on minor nuances and details.

This paper describes a test with 3 participants, this is the size of the design team in the cooperating mobility company. More companies might benefit from semantic alignment within their design team, however, the bigger the group, the more difficult and inefficient the collective labelling process. It would be interesting to test the maximum group size for effective image labelling sessions for participants to align semantics.

The individual image labelling process could be the same and remain effective. Even though it becomes a repetitive task, participants understood that individual labelling is essential to facilitate a thorough discussion which leads to semantic alignment. Interestingly, on average the collective labels consist of fewer words than the individual labels. Perhaps this lies in the group's effort to concisely formulate a label with minimal but effective terms. The group's discussion does not lead to incorporating a bit over every label but combining the individual description into a specific collective label.

Combining labels in small groups allows the participants to have open discussion, when the group size increases, these discussions would not flow as naturally and speaker dominance would play a bigger role (Fay et al., 2000). Alternative methods should be discussed before co-creative image labelling can effectively be used in bigger design teams. The group's size does not only complicate time-efficient labelling but there would also be too many perceptions to consider when combining labels. Not all participants might feel heard in the discussion, therefore, semantics would only align amongst participants that actively participate.

Polanyi (2009) mentions that as humans 'we can know more than we can tell' where he describes what he calls 'tacit knowledge'. Explaining and communicating specific design reasoning is similar. Cross (1982) claims that 'what designers know about their problem-solving process remains largely tacit knowledge'. Ultimately, the LoRA models could externalise tacit knowledge to share it with new team members. Team members can "experiment" with the team's LoRA model to see how it links prompts to output.

A dataset with 12 images offers insufficient data to train a high-quality LoRA model which captures the full complexity of the participants' perception and articulation of images. If the model does capture the full complexity, it forms an externalisation of the participants' perception and articulation which could illustrate the perception and articulation of perception to new members or external parties. This would be an externalisation of tacit knowledge which is normally hard to articulate. A LoRA model could be a tool that visualises tacit knowledge through an interactive input-output experience. This could be an interesting direction for further research.

CONCLUSION

To conclude, co-creative image labelling sessions lead to improvement in semantic alignment within design teams. Throughout the process the labelling speeds up. Although the individual labelling feels repetitive, it facilitates sharing unbiased individual perceptions and minimises the influence of social relations. Due to the importance of discussing the labels, there could be a limit to the size of the design teams for which this method works.

This paper illustrates an promising use of AI as a mirror or externalising tool for tacit knowledge through which the users learn to understand themselves and each other more thoroughly.

ACKNOWLEDGEMENTS

We thank Stef de Groot for allowing us to research this topic at the Dutch mobility company. We also want to thank Erik Jan Hultink, Pieter Jan Stappers, Lena Hegemann, and Ricardo Mejia Sarmiento, who helped us consider new directions and different approaches to this topic. Lastly, we thank Carlo van der Valk who was an indispensable support when using Stable Diffusion and testing the LoRA models.

FURTHER RESEARCH

First, this chapter discusses two experiments I did after the image labelling session. Then it suggests several other research directions: a more elaborate single-designer image labelling session to test the externalisation of tacit knowledge, and a label-to-sketch experiment to test the effect of semantic alignment. The first test illustrates a clear improvement compared to the LoRA model trained using the group labelling session dataset. It does however also give some pointers for more accurate labelling and training. The second experiment illustrates the same problem designers faced when they used image-generating AI, a generated image can fit the prompt, but the prompt can still be interpreted in very different ways.

SINGLE-DESIGNER LABELLING FOR HIGH-QUALITY LORA OUTPUT

After the co-creative LoRA labelling session, the designers were more aligned than before, however, they were also especially enthusiastic about the thought that this could allow them to externalise their perception and interpretation for new team members to learn. I had mentioned before that this was not the primary goal of the session, but the designers wanted to know if, when they would label more images, it would be possible to train the model and get desirable generated results. They wanted to know if the model would generate images that fit the prompt like they generated labels that fit the training images.

Additionally, the designers were curious about how well the LoRA model would understand the labels if the training dataset contained more images that were applicable to the Gazelle design style. Then the model would have to pick up more nuances and smaller details and differences in the training dataset.

Using the label structure formed during the co-creative image labelling session, a single designer tried to label as many images as possible in two hours whilst maintaining a consistent labelling format. He labelled 23 form studies in total, all by Kobayashi (Kobayashi, 2013; Kobayashi, 2017). He focused on using the same terms and structure in his labels to maximise the LoRA's performance.

I had preselected images that were closer to the Gazelle design language and product design in general. During the session, the designer mainly went from top to bottom, but skipped some images because they were too vague, or conflicting, or because the shapes did not have enough "body". During the last few minutes, he picked some remaining images that had under-represented or unique features that he thought were interesting to include. While labelling the first few images, he spent a lot of time going back and forth adding labels he had forgotten, and removing labels that he decided were abundant. Therefore, the labelling did not necessarily speed up significantly.

Because the dataset contained more similar pictures, the designer had a harder time labelling images that fit the exact nuance of the aesthetic he was describing. He increased the use of "semi" and chose to leave subjective terms out several times because the shapes were in the middle ground between, for example, organic and geometric.

The designer also explained that there was a lot of reasoning behind the labels he was giving. The reasoning helped him give consistent labels, but he could not share the reasoning with the LoRA model. Throughout the session, the designer once again realised the difficulty of accurately describing forms. He also felt that some definitions changed as he saw and compared more images.

After labelling the images, we trained the LoRA model and evaluated the output. The designers recognised some labels better than others. Some results are below:

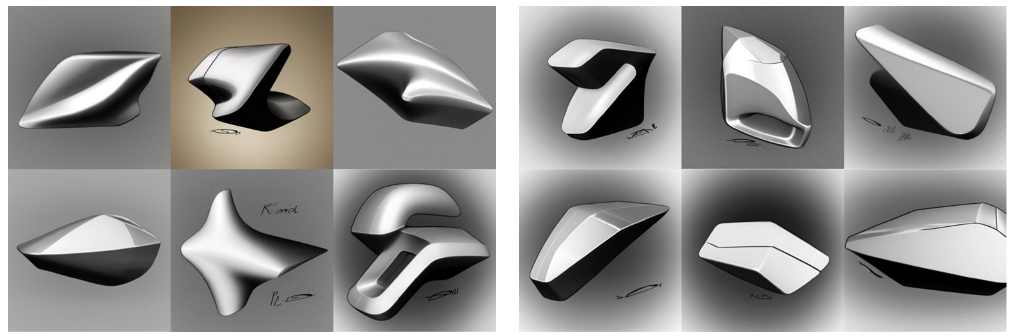


Figure 4.1. Stable Diffusion output using the LoRA model trained on the single-designer labelled dataset. The images on the left were generated with the term "organic", the images on the right were generated with the term "geometric".

The labels "organic" (left) and "geometric" (right) came out quite strongly. These are labels that stem from individual visual features as angular or curved lines have a big influence on the perception.

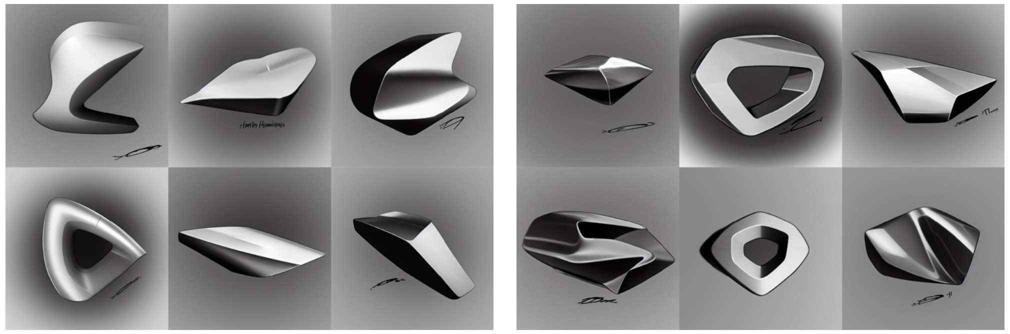


Figure 4.2. Stable Diffusion output using the LoRA model trained on the single-designer labelled dataset. The images on the left were generated with the term "harmonious", the images on the right were generated with the term "chaotic".

The labels "harmonious" (left) and "chaotic" (right) are slightly less recognisable because these depend on the combination and interactions between different visual aspects.

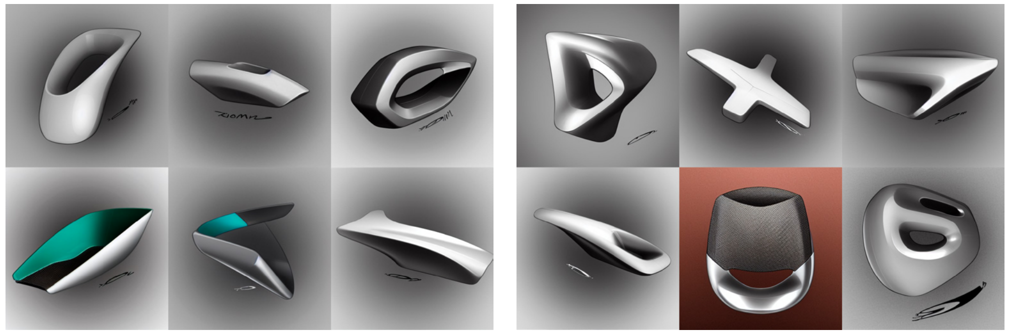


Figure 4.3. Stable Diffusion output using the LoRA model trained on the single-designer labelled dataset. The images on the left were generated with the term "elegant", the images on the right were generated with the term "powerful".

The labels "elegant" (left) and "powerful" (right) came out debatably well. There are some elegant features but those are not necessarily the opposite of powerful. Perhaps these terms are not proper antonyms.

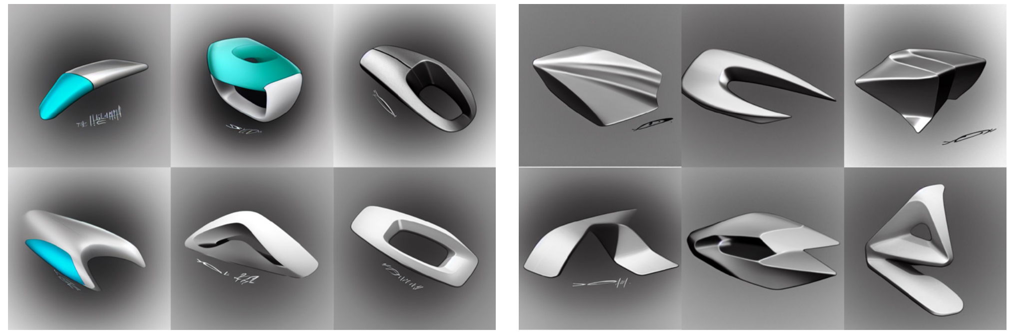


Figure 4.4. Stable Diffusion output using the LoRA model trained on the single-designer labelled dataset. The images on the left were generated with the term "minimalistic", the images on the right were generated with the term "complex".

The labels "minimalistic" (left) and "complex" (right) translate very well. Most minimalistic shapes have no unnecessary busy features. The complex shapes do have odd or unexpected shapes.

This dataset also contained several images with coloured parts, these strongly translated to several labels. It makes the generated images a bit more colourful, but perhaps, for more objective labelling, the pictures should be greyscale.

Figure 4.5. Stable Diffusion output using the LoRA model trained on the single-designer labelled dataset. The images on the left were generated with the term "flowing edges", the images on the right were generated with the term "continuous edges".

The images show promising output on the higher level labels, but details like "flowing" vs "continuous" edges do not always translate as well. Both also seem to link to labels like organic and geometric.

The results are promising for high-level aesthetic terms. One could design a method so new team members quickly learn these terms and learn to understand the mental model of a design team through interacting with a LoRA model. This would be similar to how prompt engineers learn to understand the mental model of generative AI like DALL-E and Midjourney. One could "vibe" with the AI, which should also indicate vibing with the team.

POST SESSION LABEL TO SKETCH

During the labelling session, participants write labels for images. The idea is that they would be more aligned in their designs when agreeing on textually defined or discussed design language. To validate the effectiveness of the labelling session, I wanted to experiment with going from label to image. I asked two participants to turn three silhouettes with labels into sketches. I then checked how comparable the results were. The results can be seen in Figure 4.6.

The first two silhouettes are irregular and unrecognisable shapes, and the third shape is a front tube joint which is a component both participants have designed before.

When looking at the first two shapes, the participants' visualisations. Both participants translate all textual elements into their visualisations, however, the shapes are different. They interpret the "sharp flowing edges" similarly, but the volumes for both drawings are very different. The extent to which the edges flow also differs, participant 1 has some lines that feel slightly like softened corners, and participant 2 visualised almost only curved edges. For the second silhouette, both shapes are also different. One shape is more of a shell, while the other seems very voluminous. Both shapes have a chamfered edge, flat surfaces, appear geometric, and have sharp edges. This illustrates that the silhouette with a description with semantically aligned terminology cannot fully control how the visualisation turns out.

The last silhouettes are more similar because the description specifically mentions "front tube joint" which both participants are familiar with. This causes their volumes to match more closely. Additionally, the textual elements can also be recognised in the visualisations. The visualisations show the same sharp flowing edges and similar harmony. This hints that the description could help the detailing in a similar design language for defined volumes. These two volumes are a lot closer to a similar design family than the other sets of shapes.

These findings indicate that a description does not fully control the outcome of visualisations, but based on the given requirements, it can guide the detailing in similar directions.

As mentioned when analysing the image-generating AI, "a picture is worth a thousand words" so words will rarely give complete control of image generation, this is also true when humans create the image. During a classic design process, the prompt can be explored and discussed in more detail

before finalising a design. This offers more control points to better explain and understand the desired output.

Form study of an organic powerful and harmonious shape with sharp flowing edges and a small cavity, soft surfaces and a soft silhouette	Form study of a soft geometric harmonious and elegant shape with chamfered sharp geometric edges, straight angular split lines and flat surfaces, sharp silhouette	Form study of a front tube joint with soft geometric elegant harmonious shape, soft curved surfaces, sharp flowing edges, soft geometric silhouette

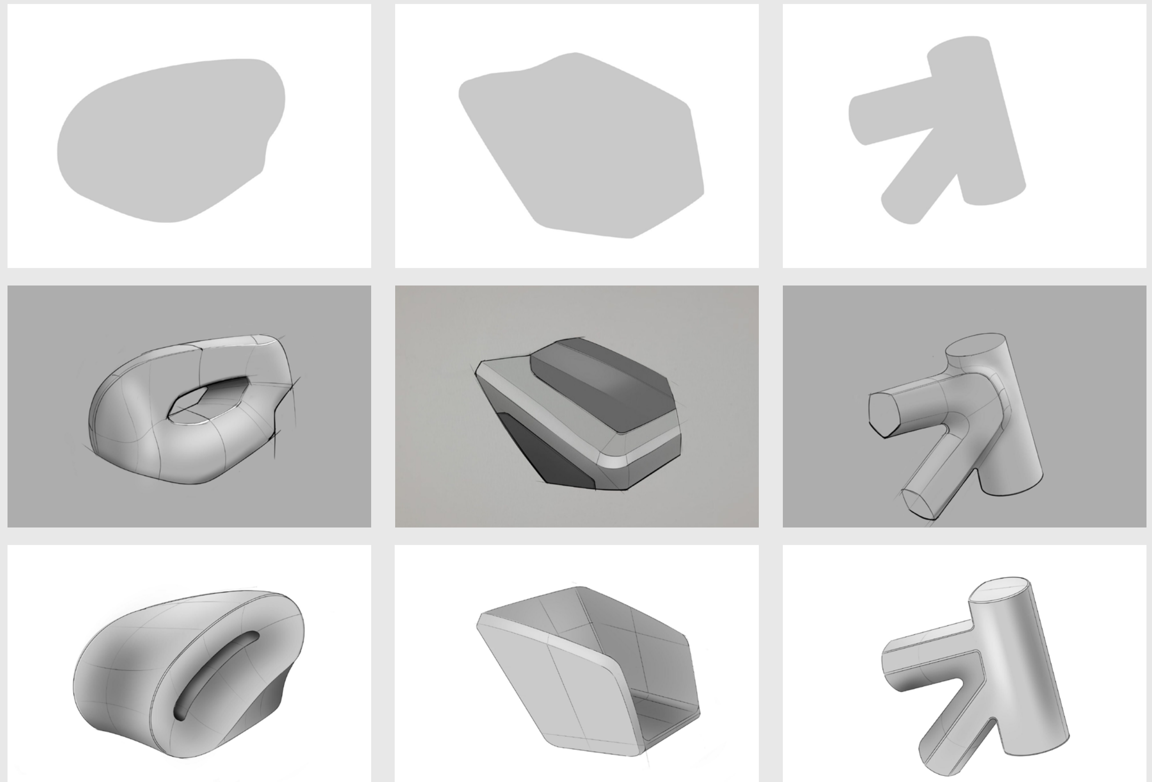


Figure 4.6. Each column shows a given description and silhouette followed by two visualisations created by two designers at Gazelle.

The semantic alignment does appear to lead to similar detailing which can help align aesthetics. This is similar to the LoRA model in the previous experiment that captured some, but not all textual aspects. Labels do not fully define a shape giving designers sufficient design freedom to use their creative freedom and generate unique and aesthetically pleasing designs but also giving sufficient guidance through the frame captured in the description.

FURTHER RESEARCH POSSIBILITIES

Apart from the two experiments above, several other interesting directions can be discovered in the future. Firstly, this method should be tested with multiple test groups and different group sizes. During another test, a group could focus more on detailing nuance-specific or very detailed features. Furthermore, designers could see if the generated output effectively works as inspiration. Focusing on the model, one could test LoRA models as externalisations of a group's vibe, especially how the externalisation could be used. Last, it could be interesting to apply co-creative labelling for different creative industries, like music, fine arts, or theatre. This section briefly goes into these proposals.

CO-CREATIVE IMAGE LABELLING VALIDATION

Currently, the effects of co-creative LoRA model image labelling for semantic alignment have only been tested in a group of three. I could not determine if the effects were significant. If the session were executed with several design teams, the effects could be measured more thoroughly and a better claim could be made on the effect of the session.

CO-CREATIVE IMAGE LABELLING GROUP DIMENSIONS

During my test, the group consisted of three participants. This led to a natural discussion where all participants could talk. However, many design teams will have different team sizes. Experiments could be done with a minimum, probably two participants, and a maximum group size for this session to be effective and efficient. In bigger groups, social dominance can quickly play a big role.

CO-CREATIVE NUANCE LABELLING

Currently, the dataset labelled by the participants consists of a wide variety of images within the same style. The dataset labelled by the single designer was more specifically focused on product design shapes the designer marked as relevant. If a dataset becomes even more specific during a co-creative labelling session, it could facilitate a discussion on very detailed and nuanced terminology and accurately align the vibe within a participant group. It could also backfire in a discussion where participants cannot agree or cannot articulate their perceptions at all. This would be an interesting direction to explore.

GENERATED FORM STUDIES AS INSPIRATION

The LoRA model trained on the self-labelled dataset can generate form studies fitting the designer's or design team's semantics. Therefore it should generate examples that fit the textual prompt they give. The generated output could be inspirational when designing a product in a textually specified style. Designers could generate several examples of geometric and powerful shapes, and use these as inspiration. It could be interesting to see how well (artificial) form studies function as inspiration. Normally the explorative process of creating form studies is most inspirational, perhaps the images alone can also be inspirational.

USING LORA MODELS AS EXTERNALISATIONS OF TEAMS' VIBES

The LoRA model is an artificial representation of how designers articulate their perception of images. Using the model within Stable Diffusion enables users to reverse the process and generate images based on the labels. Both text-to-image and image-to-text could be ways to share the group's vibe with new team members or designers. Usually, when employees leave a company, they leave a knowledge gap. Companies minimise this gap through a transition period when new employees start, and through proper documentation. Tacit design knowledge has always been difficult to report or quickly transfer to new employees. The LoRA model is an externalisation of such tacit knowledge, the question remains how to best use such an externalisation, and how could it transfer tacit design knowledge? I believe this would be an incredibly valuable and interesting research direction.

CO-CREATIVE LABELLING FOR ALTERNATIVE INDUSTRIES

Last, there are more industries in which definitions are ill-defined. For example, music genres are not clear-cut but transition into each other. Co-creative labelling might also help better define communication about musical genres and music moods. This could facilitate more effective communication between artists or between artists and producers to efficiently work towards their goals.

The risk for all of these directions is that co-creative labelling might lead to the conclusion that current vocabulary does not support the nuance or detail required to accurately discuss certain topics. Therefore, people might create more sector-specific labels or lingo, increasing the threshold for new entrants to easily enter the team or sector.

DISCUSSION

This report touches upon numerous experiments, thoughts, and insights related to image-based AI in the industrial design environment. During this project, I had the luxury of discovering this emerging landscape with a real-world industrial design company producing products that balance a diverse set of requirements and desires, from aesthetics and ergonomics to producibility and integration of technology.

This discussion briefly describes some general thoughts on the current state of image-based AI for industrial design, the bigger picture of co-creative image labelling for semantic alignment, and the future of AI in the industrial design environment.

CURRENT STATE OF AI FOR INDUSTRIAL DESIGN

Image-based, especially image-generating AI has rapidly gained popularity over the last few years. Curious designers have experimented with AI to see what it can do. Usually, experiments start with amazement, being surprised by the image quality and diversity. Depending on what they try to do, amazement turns into excitement or critique. When they simply want “a picture” they are excited with the output. When people want a picture of something specific, they can end up disappointed because the AI does not always generate the desired output. The more specified their input, the bigger the chance the output disappoints. This is what the crooked bowtie effect describes (Verheijden & Funk, 2023).

Regardless of its limitations, I advise all designers to experiment with AI. Tools with limitations can still add value to the industrial design process. Awareness of these limitations prevents unnecessary frustrations and facilitates more effective use. No one knows how to use these tools best (if there even is an optimum), so experiment freely. Similar to a designer modelling in CAD software for the first time, part of using AI tools depends on the skill one has. If one knows how to write good prompts, good output follows.

When designers experiment with and know the basics of a wide set of tools, they can use those that fit their task best. Each tool has its limitations. For example, Midjourney does not offer control to maintain or edit specific details, therefore ControlNet could be a better fit. Similarly, Solidworks has a functional rendering feature, but Keyshot performs better at that specific task.

Training AI models takes time but can help in the long run. Designers can benefit from quick image generation and a strong vibe between the designer and AI helps the generative process. Designers can train models to strengthen the vibe and steer generation towards a specific domain, style, or product category. This can make AI output more relevant to the designer’s needs.

When experimenting with AI tools, be open to the output and save a critical mindset for later stages. I have a few tips for implementing AI in the industrial design process. By no means are these tips objective or academically grounded, these are simply the things I noticed during this project.

1. Be open to undesired or unexpected output. You do not have full control, if you expect it, frustrations arise.
2. Good looks do not equal good reasoning. AI can generate high-quality pictures that would take humans hours to produce, but it does not have a reason for how these pictures came to be.
3. AI can only generate output based on training data. Training data consists of images and labels created by humans. Many creatives doubt the capabilities of AI to create truly unique and innovative output.
4. Not all AI tools seamlessly integrate into the design process. Using tools takes time you cannot invest in other steps. Tools like KREA function on top of the current process without obstructing it, Vizcom is half-half, but text-to-image generators require people to step out of the process. Check you can spend the time on the detour.
5. AI tools require skill, do not expect to master AI tools within seconds, practice, practice, practice!

CO-CREATIVE IMAGE LABELLING FOR SEMANTIC ALIGNMENT

The image labelling session focuses on labelling images for a LoRA model. However, the LoRA model does not directly add to the aligning functionality of the session, solely co-creative labelling would suffice. Then why does this method link to AI? This is a question I asked myself countless times during my project. I came to two key conclusions: AI training motivates the team to label accurately, and designers and AI need these sessions to progress together.

When labelling the image dataset, using the dataset to train an AI model motivates the participants to label accurately. The participants reflected that the labelling itself felt like doing homework. Labelling felt like an indirect approach to a clear problem. The idea that the dataset would be used to train a model created an additional goal: to train an AI model that understood the participants' perceptions and articulation of images; thus their shared vibe. They were motivated to accurately describe and discuss their perception, otherwise the model would not recognise the patterns. Without this goal, motivation to label the images seems pointless because the labels would not be used for anything. The training of AI motivates the designers to align among themselves.

For AI in the industrial design realm to progress to the next level, designers need to take an active role in the development of AI. As mentioned several times in this report, current AI models do not understand the terminology and specificity with which designers articulate their ideas and perceptions. Regardless, designers want to use AI tools but they quickly meet the misalignment as an obstacle. Before AI models can be trained to understand designers better, AI needs datasets with design-specific labels for training.

Labelling sessions could facilitate the creation of these design-specific datasets but they do need some improvements. Firstly, participants said it was quite boring. Making it more fun would motivate people to create these datasets. Additionally, labelling specific aspects of imagery could strengthen the understanding and patterns recognised by AI models. Van der Burg et al. (2023) did not write labels for entire images but for specific features in images. This could possibly improve the model quality.

Additionally, the aligning effects of the co-creative labelling session should last longer. After four weeks, the alignment was slightly worse than before the session. Participants still knew some of the terminology, but they did not always use it for the same aspects. Their articulation of perception had misaligned. It would be valuable to develop a method where the alignment would last longer, possibly by collectively labelling one image every other week. Another option would be to actively encourage the use of specific terms in day-to-day work. Ideally, this happens naturally, if it does not, perhaps the images in the dataset should illustrate different objects to make the discussions more relevant.

The motivator for participants, training an AI model on the labelled dataset, could be used as an externalisation of the participant group's vibe. It would hence be an externalisation of tacit design knowledge, something that is normally difficult to communicate and shared through collaboration. One could compare models over the years and see how the vibe changes, but also maintain a vibe when a key designer leaves the design team. The externalisation of tacit knowledge has clear value, but the interaction with such a model to effectively learn tacit design knowledge would have to be explored.

Figure 5.1. Two images generated by GPT-4 for the prompt "Could you generate an image of: electric bicycle with an integrated battery in the downtube, an external headlight and an active frame"

FUTURE OF AI FOR INDUSTRIAL DESIGN

Predicting the future is guesswork but desiring a future is a basic human skill. I quickly want to mention some directions I hope the future of AI for industrial design will develop towards.

One key direction is a designer's ability to read and write code. This research explores many AI models and tools which were all coded by someone else, luckily, I could edit their parameters through no code interfaces. Some of these changes I would have been able to, but I do not have the skill to quickly alter major functionalities. Understanding some of the code is helpful, but as a designer, we will rarely outperform a computer scientist. I recommend designers learn some basic code and understand the functionality of AI so they can converse with coders and explain their needs. This is the most important value designers can add, thoroughly explaining what they need for the tools to work, usually, others are more equipped to create them. But what is it we would like to see?

I would like to generate output with reasoning. I could tell current models to generate an "electric bicycle with an integrated battery in the downtube, an external headlight and an active frame" (Figure 5.1). It might come up with one or more designs and I would like or dislike features. Perhaps I would prompt "create a different profile for the frame" for which it generates output. However, if I had a similar conversation with a colleague designer, I would say "We should create a different profile for the frame because [e.g. this is too bulky to fit out target groups' desired]. In naming the why, I am learning about the problem. The colleague would ask things like "why doesn't our target group like bulky frames?" or "the bulky frame does create the desired stiffness, which one do we think is more important?". This would facilitate an explorative process but keep the quick iterative generative process intact.

Additionally, designers spend most of their time developing their designs in CAD software. They would really like AI to facilitate a more intuitive and fast-paced workflow.

This comes down to my overarching thought. The design process is a complex and multifaceted process in which designs develop an understanding of the problem while creating a solution. I have always liked this balanced puzzle of exploring, understanding, and creating. I would not want AI to take that fun away from me, but I would love it if AI could minimise my boring tasks, speed up my insights, and broaden the problem and solutions spaces I can discover. Help is always welcome, but I would not want AI to take over the fun job of designing. Even if it could eventually, I do not think designers and consumers will allow it. Especially in high-end design, we value design because it was created by people.

CONCLUSION

To map the opportunities and limitations of image-based AI, this thesis explores and tests several tools for and with designers at Royal Gazelle. The solution-finding phase highlights two key topics: the design process at Gazelle, and the general opportunities and limitations of image-based AI tools. Researching these topics led to the identification of misaligned perception and articulation of perception among designers as a key problem to solve. During the solution-findings phase, co-creative image labelling for LoRA models was estimated as the best solution. This phase was spent preparing for this session and evaluating the results afterwards. The paper summarises the key considerations and conclusions from the session. Lastly, further research describes how co-creative image labelling needs additional verification but can be developed further for more insights.

Designers at Gazelle translate a mainly textual proposition into a 3D design which they hand over to engineering. They combine numerous aspects in their design like aesthetics, ergonomics, user experience, producibility, etc., but overall, their biggest challenge is to create a coherent product portfolio with recognisable aesthetics.

When using image-based AI tools, designers recognise how it can speed up and broaden the design process. It enables them to think of ideas they would normally never have. However, as their explorative process becomes more demanding, the designers desire more control over the tools to generate relevant images. The lack of control stems from two key things: the minimal ability to textually explain images in full detail, and misaligned semantics between AI and designers.

The second cause for the lack of control over AI tools relates to the challenge identified in the design process. Misaligned semantics between AI and designer stems from misaligned semantics among humans. Semantic misalignment among humans complicates designing a coherent product portfolio in teams because they cannot discuss their designs effectively. Since semantic misalignment is present among humans, it is also present in AI training datasets and between AI and designers. In this research, designers co-creatively label image datasets for LoRA models to align semantics among themselves and between them and AI.

The best type of images to label was determined in preparation for the labelling session. Form studies are the least ambiguous category, yet portray sufficient features for a relevant discussion. Images label the images in sets of five, first, they write five individual labels, then they write the collective label, and then they move to the next five. This provides a balance between individual perception and input, and aligning perception and articulation during the session. Participant alignment is measured using a quantitative pre- and post-session questionnaire, a reflection questionnaire, a verbal reflection, and by analysing participants' scratch papers, spreadsheets, and audio recordings from their discussion. After four weeks, participants complete the quantitative questionnaire again for further research.

The labelling session increased the perceptive alignment by 30.8%, although this is a noteworthy difference, it is not statistically significant due to the small sample size. Participants do use more similar vocabulary after the session compared to before. Lastly, all participants reflected that the session helped them understand how they interpret form studies themselves, how others do, and how the group does. Additionally, the group felt that there was a better collective understanding of how to describe form studies. After four weeks, the participant group had lost alignment. They did remember (parts of) the vocabulary but participants used it differently.

After the session, the dataset was used to train a LoRA model and externalise the participant group's vibe. The model captured some features but lacked nuance. To test the possibility of externalising a vibe, a single designer labelled a bigger dataset and trained a LoRA model to generate several images. Some of the images fit their prompt, but some of the output lacked some details. The LoRA models can potentially be an externalisation of a design team's vibe but do not currently capture the full complexity.

Semantically aligning the participant group does affect the creative process of the participants. When two participants were given a textual description and a silhouette, both designers created 2D visualisations of 3D objects. Although the general shapes differed, their detailing matched and the textual descriptions were translated similarly. When participants were tasked to design a known object, a front tube joint, based on a textual prompt, they created even more similar visualisations.

In further research, the co-creative labelling session should be repeated with multiple participant groups to validate the significance of the effect on semantic alignment. If the LoRA would be used as an externalisation, it would be interesting to research how such an externalisation could best be used to share the vibe. Lastly, the labelling session might also be useful to other industries.

This report illustrates that AI does not only help us and provide us with information but interacting with or training AI models can give humans new perspectives and skills.

REFLECTION

Over the last couple of years, many new AI tools have been launched, some of which have made the news and have become immensely popular. This development continued through my graduation project. New models were launched, models were updated, and more and more research was published on the intersection between AI and design.

At the beginning of the project, I was asked to look into possible use cases of AI for designers at Koninklijke Gazelle. The research question was so broad that I realised I needed some reframing and strict deadlines to deliver something of value without losing myself in research. I reframed the scope of the research to specify “image-based AI”.

I started my project with two parallel research topics. On the one hand, I tried to understand the design and product development process at Gazelle. I did this to try and map the opportunities where the design could use help. On the other hand, I looked at and experimented with several AI tools that seemed fitting for the design practice. I did some of the experiments on my own after which I validated findings with the designers, I did other experiments together with a designer at Gazelle to see how he reacted to the functionalities. Afterwards, I compared the findings from both research topics and found several possible combinations. Based on conversations with the designers, innovation manager, and my supervisory team, I chose to develop an image labelling session which would facilitate the semantic alignment between designers.

After choosing that direction, I had to learn to work with image labelling, LoRA models, and how to set up a session that would facilitate the right discussions. I did some theoretical research, preparatory thinking, and a pilot to test the session’s effectiveness. The real session worked well, both the qualitative and quantitative analysis indicated that the session was useful.

In the end, I developed a methodology that proved an alternative approach to AI. AI is not only something that we can use to take work off our hands, it is also something that can teach us to evaluate our own workflows and mental models, or even to understand each other and align common perceptions.

When I look back, I am mostly glad with the results. I am glad that I considered a method that could help designers and that I could prove that it works. I have mixed feelings about the research that followed, follows and the implications it has on the designer’s workflow. After the initial experiment, there were a ton of experiments that I thought would be funny, but I also realised that the designers at Gazelle weren’t readily available to do many more tests, which proved to be the difficulty of doing research in a commercial setting. I should be, or have been more assertive and demanding of their time as it would benefit them too.

REFERENCES

1. Annis, L. F. (1983). The processes and effects of peer tutoring. *Human Learning: Journal of Practical Research & Applications*.
2. AquaHug. (n.d.). *I created a program called Vizcom that uses Ai to automatically render sketches. : r/Design*. https://www.reddit.com/r/Design/comments/on670l/i_created_a_program_called_vizcom_that_uses_ai_to/
3. Arcas, B. a. Y. (2016, May). H*ow computers are learning to be creative [Video]. TED Talks*. https://www.ted.com/talks/blaise_aguera_y_arcas_how_computers_are_learning_to_be_creative?language=en#t-331693
4. Automotive Design Planet. (2021, October 10). *VIZCOM: THE AI FOR ARTISTS AND DESIGNERS | LinkedIn*. https://www.linkedin.com/pulse/vizcom-ai-artists-designers-automotive-design-planet/
5. Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., ... & Ramesh, A. (2023). Improving image generation with better captions. *Computer Science*. https://cdn. openai. com/papers/dall-e-3. pdf, 2(3), 8.
6. Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2), 115.
7. Bieler, G. (2024, February 28). A comprehensive guide to training a Stable Diffusion XL LORA: optimal settings, dataset building, captioning and model evaluation. *Medium*. https://medium.com/@guillaume.bieler/a-comprehensive-guide-to-training-a-stable-diffusion-xl-lora-optimal-settings-dataset-building-844113a6d5b3
8. Booth, B., Donohew, J., Wlezien, C., & Wu, W. (2024, March 5). *Generative AI fuels creative physical product design but is no magic wand*. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/generative-ai-fuels-creative-physical-product-design-but-is-no-magic-wand
9. Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *The Journal of the Learning Sciences*, 6(4), 347-365.
10. Cross, N. (1982). Designerly ways of knowing. *Design studies*, 3(4), 221-227.
11. Davenport, T. H. (2022, November 2022). *How generative AI is changing creative work*. Harvard Business Review. https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work
12. Dorst, K. (2006). Design problems and design paradoxes. *Design issues*, 22(3), 4-17.
13. Duran, D. (2017). Learning-by-teaching. Evidence and implications as a pedagogical mechanism. *Innovations in education and teaching international*, 54(5), 476-484.
14. Ellis*, R. A., Calvo, R., Levy, D., & Tan, K. (2004). Learning through discussions. *Higher Education Research & Development*, 23(1), 73-93.
15. Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological science*, 11(6), 481-486.
16. Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene?. *Journal of vision*, 7(1), 10-10.
17. Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, 38(4), 281-288.
18. Gilchrist, A. (2012). Objective and subjective sides of perception. *Visual experience: Sensation, cognition, and constancy*, 105-26.
19. Goldschmidt, G. (1991). The dialectics of sketching. *Creativity research journal*, 4(2), 123-143.
20. *Google Sheets*. (n.d.). Google Sheets. https://www.google.com/sheets/about/
21. *GPT-4*. (n.d.). https://openai.com/research/gpt-4
22. *Gravity Sketch*. (n.d.). Gravity Sketch. https://www.gravitysketch.com/
23. Gu, Y., Wang, X., Wu, J. Z., Shi, Y., Chen, Y., Fan, Z., ... & Shou, M. Z. (2024). Mix-of-show: Decentralized low-rank adaptation for multi-concept customization of diffusion models. *Advances in Neural Information Processing Systems*, 36.

24. Guinness, H. (2024, February 22). *The best AI image generators in 2024*. Zapier. https://zapier.com/blog/best-ai-image-generator/

25. Hollowstrawberry. (2023). *kohya-colab/Lora_Trainer.ipynb at main · hollowstrawberry/kohya-colab*. GitHub. https://github.com/hollowstrawberry/kohya-colab/blob/main/Lora_Trainer.ipynb

26. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

27. Italdesign. (2017, March 21). *Giugiaro Design and Royal Dutch Gazelle create No.1 a futuristic bike*. https://www.italdesign.it/project/gazelle-no1/

28. Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3128-3137).

29. Kobayashi, K. (2019). Interactivity: A potential determinant of learning by preparing to teach and teaching. *Frontiers in psychology,* 9, 426734.

30. Kobayashi, Y. (2013). *Collection of Forms*. Coroflot. https://www.coroflot.com/yoxkobayashi/COLLECTION-OF-FORMS

31. Kobayashi, Y. (2017). *Collection of Forms #2*. Coroflot. https://www.coroflot.com/yoxkobayashi/COLLECTION-OF-FORMS-2

32. Krea AI. (2023, November 21). *Krea AI founders*. https://krea-ai.com/krea-ai-founders/

33. Kreuzbauer, R., & Malter, A. J. (2005). Embodied cognition and new product design: Changing product form to influence brand categorization. *Journal of Product Innovation Management*, 22(2), 165-176.

34. Lawton, T., Grace, K., & Ibarrola, F. J. (2023, July). When is a tool a tool? user perceptions of system agency in human–ai co-creative drawing. *In Proceedings of the 2023 ACM Designing Interactive Systems Conference* (pp. 1978-1996).

35. Li, J., Li, D., Xiong, C., & Hoi, S. (2022, June). Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. *In International conference on machine learning* (pp. 12888-12900). PMLR.

36. Li, L., Zeng, H., Yang, C., Jia, H., & Xu, D. (2024). Block-wise LoRA: Revisiting Fine-grained LoRA for Effective Personalization and Stylization in Text-to-Image Generation. *arXiv preprint arXiv:2403.07500*.

37. Li, Q., Luo, S., & Shi, Z. (2009). Fuzzy aesthetic semantics description and extraction for art image retrieval. *Computers & Mathematics with Applications,* 57(6), 1000-1009.

38. Linaqruf. (2023). *kohya-trainer/kohya-LoRA-dreambooth.ipynb at main · Linaqruf/kohya-trainer*. GitHub. https://github.com/Linaqruf/kohya-trainer/blob/main/kohya-LoRA-dreambooth.ipynb

39. *Luma AI*. (n.d.). Luma AI. https://lumalabs.ai/

40. *Microsoft Teams*. (n.d.). Microsoft.com. https://www.microsoft.com/en-us/microsoft-teams/group-chat-software

41. *Midjourney*. (n.d.). Midjourney. https://www.midjourney.com/home

42. *Miro*. (n.d.). Miro. https://miro.com/

43. *OpenArt*. (n.d.). OpenArt. https://openart.ai/

44. Park, C. W., Milberg, S., & Lawson, R. (1991). Evaluation of brand extensions: The role of product feature similarity and brand concept consistency. *Journal of consumer research*, 18(2), 185-193.

45. Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., ... & Rombach, R. (2023). Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*.

46. Polanyi, M. (2009). The tacit dimension. In *Knowledge in organisations* (pp. 135-146). Routledge

47. Pollock, P. H., Hamann, K., & Wilson, B. M. (2011). Learning through discussions: Comparing the benefits of small-group and large-class settings. *Journal of Political Science Education*, 7(1), 48-64.

48. *Procreate*. (n.d.). Procreate. https://procreate.com/

49. *PromeAI*. (n.d.). PromeAI. https://www.promeai.pro/

50. Ranscombe, C., Hicks, B., & Mullineux, G. (2012). A method for exploring similarities and visual references to brand in the appearance of mature mass-market products. *Design Studies*,

33(5), 496-520.

51. Reese, E. (2016, January 18). *Form Studies*. Behance. https://www.behance.net/gallery/32999257/Form-Studies

52. *RenderAI*. (n.d.). RenderAI. https://renderai.app/

53. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695).

54. Roose, K. (2022, October 21). A.I.-Generated art is already transforming creative work. *The New York Times*. https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs-dall-e-2.html

55. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 22500-22510).

56. Sihombing, J. (2012, November 19). *Digital Sketches 1*. Behance. https://www.behance.net/gallery/5982093/Digital-Sketches-I

57. Singh, N. (2023, November 11). *Luma AI Launches Genie: A New 3D Generative AI Model that Lets You Create 3D Objects from Text*. MarkTechPost. https://www.marktechpost.com/2023/11/10/luma-ai-launches-genie-a-new-3d-generative-ai-model-that-lets-you-create-3d-objects-from-text/

58. *SolidWorks*. (n.d.). SOLIDWORKS. https://www.solidworks.com/

59. Tholander, J., & Jonsson, M. (2023, July). Design ideation with ai-sketching, thinking and talking with Generative Machine Learning Models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (pp. 1930-1940).

60. TopHatch. (n.d.). *Concepts App • Infinite, Flexible Sketching*. ConceptsApp. https://concepts.app/en/

61. van der Burg, V., de Boer, G., Akdag Salah, A. A., Chandrasegaran, S., & Lloyd, P. (2023, July). Objective Portrait: A practice-based inquiry to explore AI as a reflective design partner. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (pp. 387-400).

62. Van Der Maden, W., Van Beek, E., Nicenboim, I., Van Der Burg, V., Kun, P., Lomas, J. D., & Kang, E. (2023, July). Towards a Design (Research) Framework with Generative AI. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference* (pp. 107-109).

63. Verheijden, M. P., & Funk, M. (2023, April). Collaborative diffusion: Boosting designerly co-creation with generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-8).

64. *Vizcom*. (n.d.). https://www.vizcom.ai/

65. Warell, A. (2006). Identity Recognition in product Design: An Approach for Design Management, proceedings of the 13th International Product Development Management Conference. *Milan: Politecnico di Milano.*

66. Wijaya, D. T., & Yeniterzi, R. (2011, October). Understanding semantic change of words over centuries. In *Proceedings of the 2011 international workshop on DETecting and Exploiting Cultural diversiTy on the social web* (pp. 35-40).

67. Xu, M. (2021). An exploratory study of AI creativity inspired by descriptive ambiguity.

68. Yoo, G. (2010). *Collection of Forms 2010*. Coroflot. https://www.coroflot.com/georgeyoo/collection-of-forms-2010

APPENDICES

A. MIDJOURNEY EXPERIMENT
B. GPT-4
C. VIZCOM EXPERIMENT
D. KREA EXPERIMENT
E. LUMA LABS EXPERIMENT
F. STABLE DIFFUSION - DREAMBOOTH
G. STABLE DIFFUSION - CONTROLNET
H. PROBLEM FINDING CHOICE
I. MOOD BOARDS
J. EXCEL IMAGE TYPE QUESTIONNAIRE
K. EXCEL PRE-POST SESSION QUESTIONNAIRE
L. EXCEL REFLECTION QUESTIONNAIRE
M. LORA OUTPUT INITIAL TEST
N. LORA OUTPUT LABELLING SESSION
O. LORA OUTPUT SINGLE DESIGNER SESSION
P. GRADUATION PROPOSAL

Scan the QR code for the appendices or click **here**

Acta est fabula, plaudite