UKR Journal of Engineering, Innovation and Technology (UKRJEIT)
Homepage: https://ukrpublisher.com/ukrjeit/
Email: submit.ukrpublisher@gmail.com
Volume: 1, Issue: 1 (July-December 2026) **ISSN: XXXX-XXXX (Online)**

Is it Possible for GenAI tools to Generate UI?

Nikechukwum Chinwike Ene

Keele University, Staffordshire, England

***Corresponding Author**: Nikechukwum Chinwike Ene
DOI:

Article History
Original Research Article
Received: 20-05-2026
Accepted: 25-06-2026
Published: 21-07-2026

Citation: Ene, N. C. (2026). *Is it possible for GenAI tools to generate UI? UKR Journal of Engineering, Innovation and Technology, 1*(1), 17-31.

Abstract

While Generative Artificial Intelligence (GenAI) has shown great advancement in various aspects such as text, image, and video generation, its application in user interface (UI) design remains relatively underexplored. This study examines the capabilities of GenAI tools to produce UIs that meet established design standards. An experimental framework was employed, consisting of six structured tests across four GenAI systems — ChatGPT, Microsoft CoPilot, Google Gemini, and Google Stitch — covering image-based and markup-based generations of UIs. The outputs were evaluated using criteria including visual fidelity, structure, markup code quality, and trueness to the prompts. The findings show a divide between visual ideation and functional implementation. Although generalist GenAI models can generate aesthetic visual mockups, they fail to produce production-quality markup code. Only the specialised tool, Google Stitch, demonstrated the capacity to consistently generate high-quality UIs across most tests. This suggests that reliable and consistent UI generations are currently better achievable through specialised systems. The research concludes that GenAI can generate UIs. However, they serve more effectively as a creative assistant for designers than a replacement for developers as far as UI design implementation goes.

Keywords: *Generative AI, User Interface Design, AI-Assisted Design, Human-Computer Interaction.*

INTRODUCTION

The field of user interface (UI) design is a core aspect of software development. It has traditionally been a human-centric discipline, blending psychology, information architecture, and graphic design to create digital environments. The quality of a user interface can directly affect a product's functionality and in turn render it unsuitable for use. Like many fields of human endeavour, UI design is now at an intersection with a significant technological disruptor: Generative Artificial Intelligence (GenAI). Fuelled by rapid advancements, GenAI has demonstrated an extraordinary ability to produce novel content — including text, graphics, and complex programming code — from simple natural language prompts.

The promise GenAI brings is compelling. With popular tools like ChatGPT, CoPilot, and Gemini, visual UI mockups and HTML markup can be prompted. This splits the potential application of GenAI in the field of UI design into two distinct use cases: one for the UI/UX designer (ideation) and one for the front-end developer (code generation/implementation). However, there is a significant gap between this potential application and its proven practical reliability. Key questions remain regarding the ability of these GenAI systems to adhere to complex constraints, while still producing consistent quality outputs (Shneiderman, 2020). This leads to the core research question in this dissertation: Is it possible for GenAI tools to generate UI and if so, to what extent?

Aims

The aim of this study is to investigate whether generative AI can effectively generate user interfaces that are not only functional but also meet usability and accessibility standards.

Objectives

The study is guided by the following objectives:

1. To review existing literature on generative AI and its applications in creative and design

domains, with a focus on user interface development.

2. To explore and evaluate the capability of web-based generative AI tools in creating sample user interface designs based on predefined prompts or requirements.

LITERATURE REVIEW

Understanding User Interfaces

A user interface is the point of interaction between a user and a digital system. It comprises of visual and/or auditory elements that facilitate communication between a human and a system. UIs are essential for proper human-computer interaction (HCI) and if done right can effectively minimise user errors, increase productivity and ensure safety (Shneiderman, 2020).

User interfaces generally consist of various structural and functional elements such as input components (like buttons and text fields), navigation tools, and information elements (notifications, status bars, tooltips) (Norman, 2013). Functional components go beyond the visual, containing logic for certain user interactions and feedback mechanisms. Together these elements form a system that enables users to act and interpret feedback within digital environments.

AI-Assisted Design

AI-assisted UI design refers to the integration of artificial intelligence tools and algorithms that enhance the interface design process. These systems assist in various design tasks such as wireframing, layout generation, and styling. They rely on large datasets of user interface examples to infer components, placement patterns, and usability conventions, allowing designers to focus on higher-level decision-making rather than repetitive visual adjustments. According to Marr in his article, AI design assistance also increases productivity and enhances creativity, potentially allowing for quicker project completions and lower costs for clients (Marr, 2024a).

AI-assisted design also extends to other aspects like accessibility testing and usability evaluation. For example, tools employing computer vision and natural language

processing can assess contrast ratios, predict user attention flow, and suggest adjustments for inclusivity (Lawton, 2025). Shneiderman highlights that this approach supports the concept of human-centred AI, where the designer retains creative control while AI augments the design process through pattern recognition and predictive insight (Shneiderman, 2020).

Personalised User Interfaces

AI-personalised interfaces represent a more user-centred application of artificial intelligence within UI design. The AI used in these interfaces are focused on adaptive interaction rather than design creation. They dynamically change interface content, structure, and functionality based on user preferences, behaviours, and contextual data. Good examples include adaptive recommendation engines used by platforms such as Netflix and Amazon personalise displayed items and interface layouts using analytics derived from a user's history (Wan Husain & Jantan, 2017). Similarly, intelligent dashboards and e-learning platforms modify visual emphasis or menu accessibility based on observed user expertise and frequency of use (Shneiderman, 2020).

This approach aligns with the principles of context-aware computing and HCI, where systems learn to anticipate user needs and streamline cognitive load (Norman, 2013). By utilising machine learning models, AI-personalised interfaces move beyond static design paradigms to create adaptive, data-driven experiences that evolve with the user. However, as Moran and Gibbons (2024) note, this form of personalisation remains bounded by pre-designed layout structures — the system selects from existing templates rather than generating new ones. This limitation distinguishes AI-personalised interfaces from the emerging concept of Generative UI, which seeks to dynamically construct interface elements in real time.

METHODOLOGY

Context

The dissertation uses a dual-method experiment design, splitting tests into two main categories: Image generation tests and Markup generation tests. This is so for two main reasons discussed below:

1. UI Generation Approaches: One can approach UI generation either from an image generative angle, where the AI tool creates interfaces as images, or from a text generative angle, where the interfaces are constructed from text (html markup). These are two distinct types of generative tasks and should be tested separately.

2. Professional Use Cases: An image generation test will focus on the aesthetic elements that makeup interfaces while a markup generation test will focus on the functionality of these interfaces, that is, how suitable they are during software development. The former caters to a UI/UX designer's perspective while the latter is to a frontend developer's perspective.

Six tests are performed on each of the chosen GenAI models, the first four being image generation tests and the last two being markup generation tests.

Tools and Technologies

- GenAI Tools: ChatGPT, Microsoft CoPilot, Google Gemini, Google Stitch
- Development Environment: VS Code.
- Languages: HTML, CSS.

These tools were chosen for their commonality. ChatGPT, CoPilot and Gemini are popular GenAI tools dominating the market and providing utility for various tasks for users every day, however, the fourth model (Google Stitch) was chosen for its claim to be purpose built for generating UIs. Its inclusion adds clarity on answering the question on if GenAI can truly generate UI.

Prompt Engineering

The prompts used in this dissertation gradually increase in detail, communicating more and more context to the genAI tools. They generally progress in this fashion:

- Basic – an introductory prompt to get data on what the GenAI tool can generate without much context.
- Intermediate – low context prompts that contain key deliverables but not how they come together to make up the user interface. This is to simulate prompts from less technical prompters using GenAI tools.
- Advanced – prompts with significantly greater specifications.
- Transformative – prompts coupled with an image of an already existing user interfaces aimed to manipulate them or convert them to a different format. This reference image is show below.

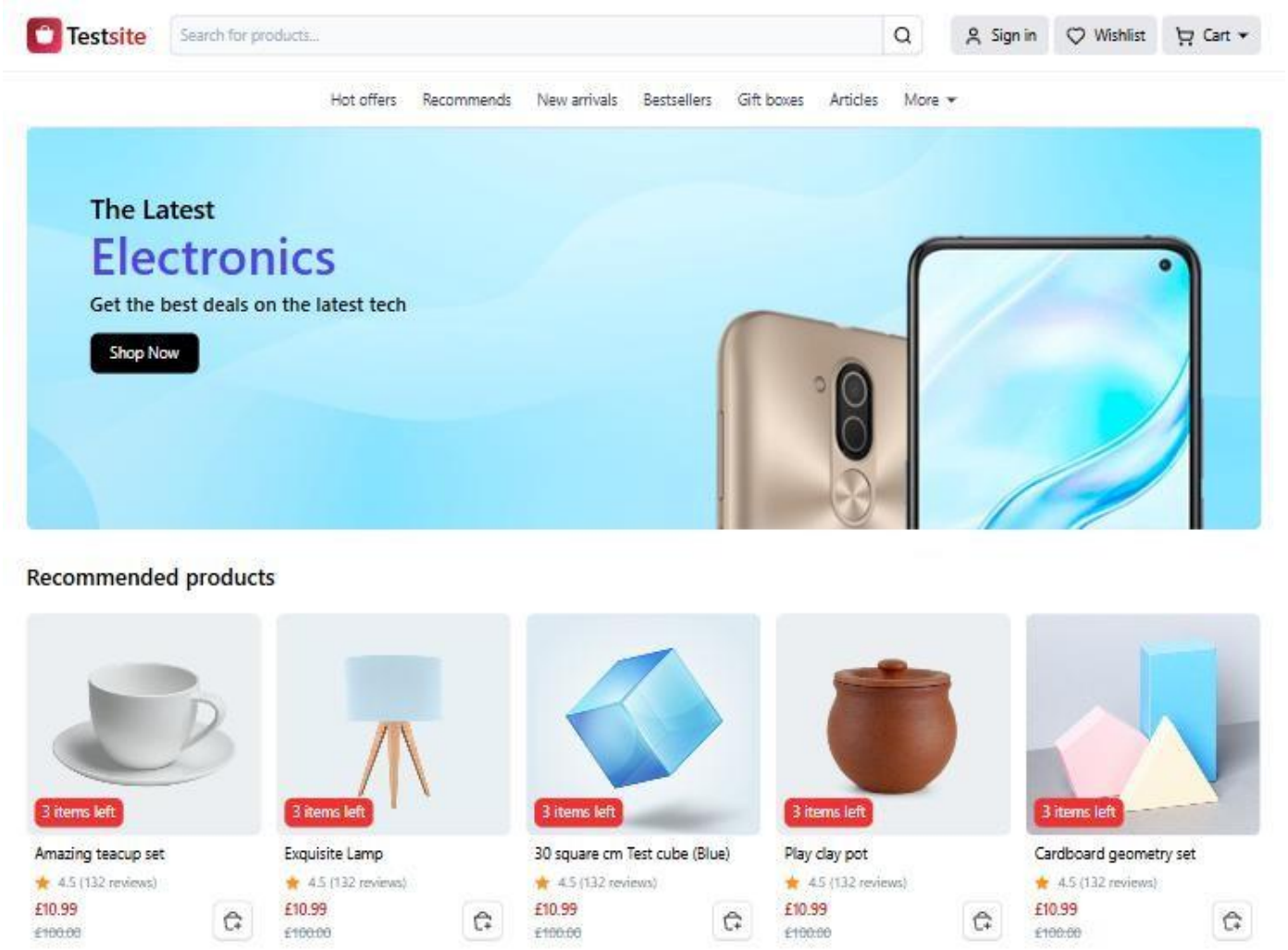


Figure 3: Screenshot of a custom-made e-commerce UI reference

Evaluation Criteria

When it comes to UI evaluations, a popular method is the usability Heuristic evaluations guidelines which have stayed relevant for decades (Nielsen, 2024). However, most of them are not easily applicable in the case of this study. For example, Neilson in his updated usability heuristic article, he mentions factors like visibility of system status, error preventions and error recovery. Since most results are static generations, custom evaluation criteria were introduced relevant to this particular case study. These

criteria streamline the tried and true heuristics into applicable factors in this study. Therefore, they are still based on factors that can affect their usefulness/reliability in both the design phases and development phases of software development. They include:

Criteria for Image Generation (Tests 1-4):

- Visual Fidelity: The overall visual design quality of the generated UI is evaluated, analysing aesthetic factors such as colour palette harmony, typography, and use of whitespace.
- Layout and Structural Coherence: This examines the logical organisation and structural integrity of the generated UI, evaluating the grouping of elements, alignment, spacing, and adherence to common UI design patterns.
- Prompt Adherence: An analysis of how accurately the GenAI output fulfils the specified instructions provided in the prompts.

3.1.2 Criteria for Markup Generation (Tests 5-6):

- Visual Fidelity: This considers how well the rendered visual output of the code aligns with the prompt's description or the source image's design, aesthetics, and layout.
- Code Quality: This evaluates the generated markup from a developer's perspective, focusing on the use of semantic HTML (e.g., <header>, <nav>), well-structured CSS, and its overall viability as a clean, maintainable starting point for development.
- Accessibility Guidelines Alignment: This criterion evaluates the basic accessibility features embedded within the generated code, focusing on factors such as the appropriate use of ARIA attributes, logical tab order for keyboard navigation, and keyboard focusability on interactive elements.
- Prompt Adherence: An analysis of how accurately the GenAI output fulfils the specified instructions provided in the prompts.

ChatGPT Tests

Test-1: Image generation – Basic prompt

Prompt:

"Generate a visual mockup of the home page user interface for an ecommerce platform"

Result:

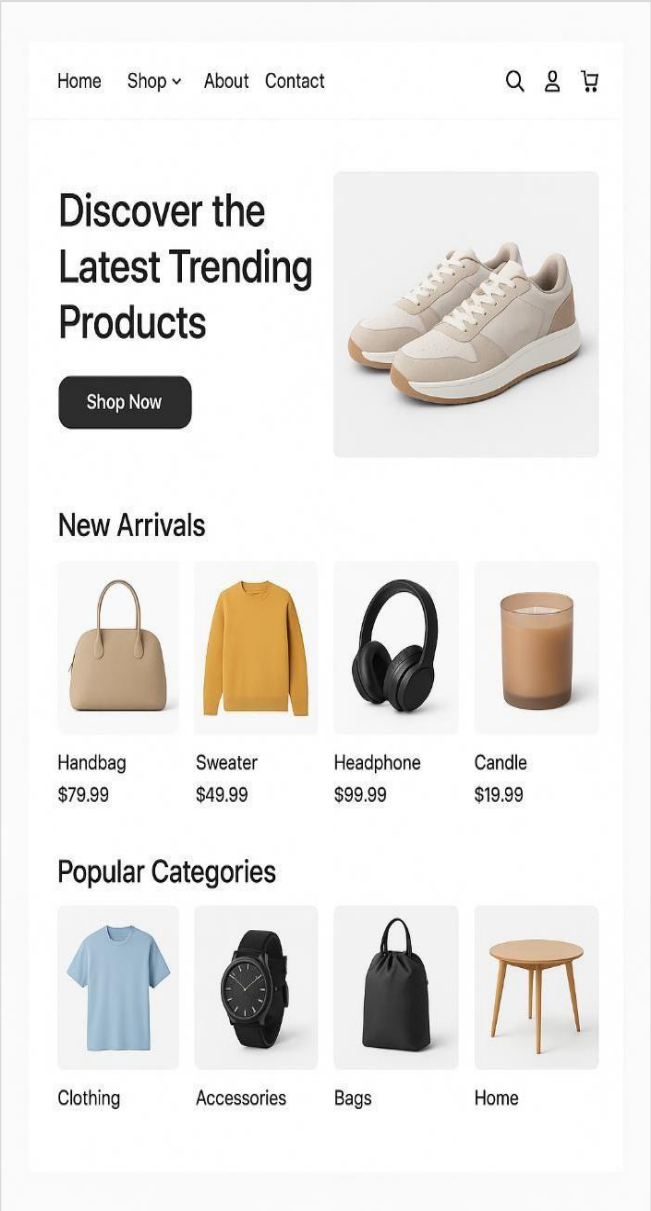


Figure 4: Generated image output from ChatGPT in test 1

2.1.1 Test-2: Image generation – Intermediate prompt

Prompt:

"Generate a mockup of a home page user interface for an e-commerce platform selling beauty products and accessories. Make the primary colour white and accent red.

Showcase various products in this home page but the main featured items are the femme fatale perfume, the cosmic flare lip gloss and the red eclipse sunglasses. Their prices are £29.99, £5.99 and £14.99 respectively."

Result:

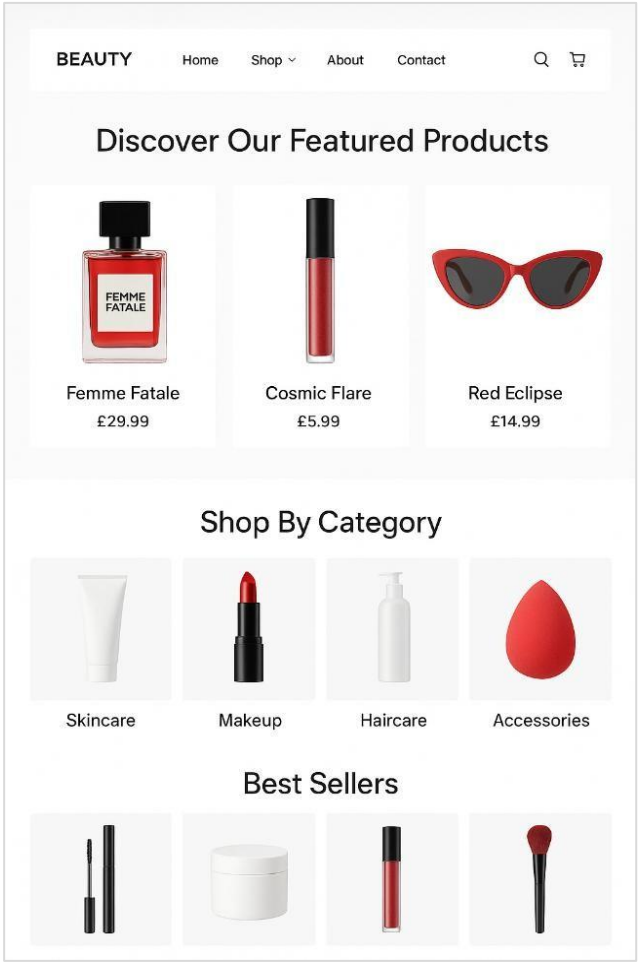


Figure 5: Generated image output from ChatGPT in test 2

Test-3: Image generation – Advanced prompt

Prompt:

"Generate a visual mockup of a homepage UI for an ecommerce platform selling beauty products and accessories. At the top, include a navigation bar with a logo by the left, category links in the middle, a search bar and shopping cart icon by the right. The hero section should display a large promotional banner with a call-to-action button with text saying 'Shop Now'. Below the hero, showcase featured products which include the femme fatale perfume, the cosmic flare lip gloss and the red eclipse sunglasses, each card including the product image, name and price (£29.99, £5.99 and £14.99 respectively). Add a section for trending items in a 4 by 4 grid, populating them with various high-quality images of accessories and beauty products. These items should all have their names and prices below their images, left aligned. The footer should include links for help, about and the social media platforms in the form of icons (specifically Facebook, Instagram and TikTok). Use the Inter font, rounded buttons, a consistent color palette consisting of white as the primary colour, black as the secondary and red as the accent colour."

Result:

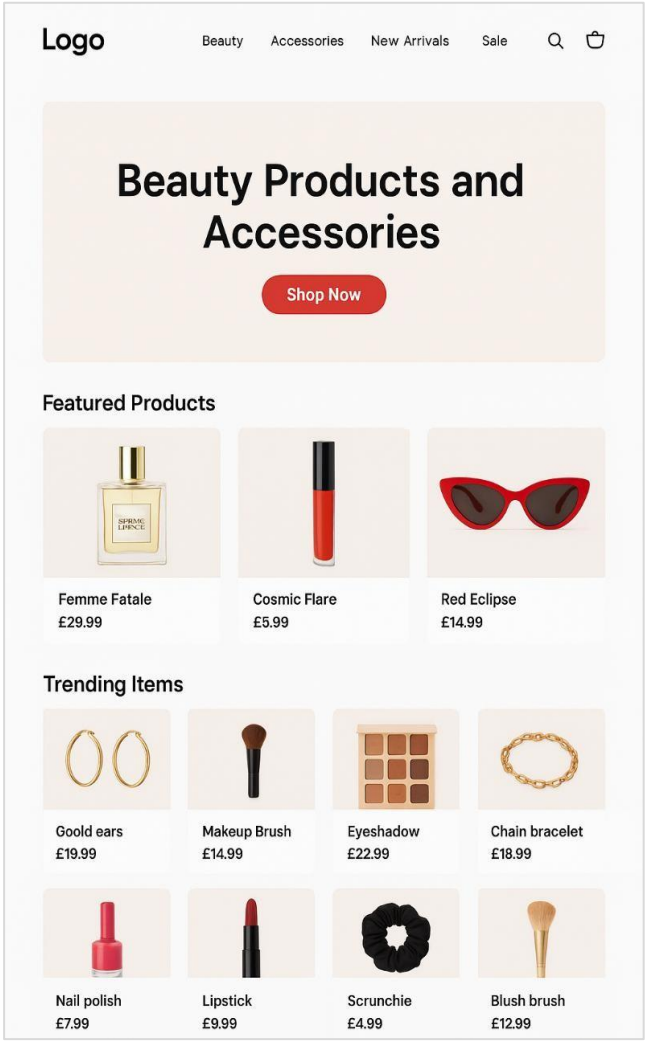


Figure 6: Generated image output from ChatGPT in test 3

Test-4: Image generation – Transformation test

Reference image used in this test is shown above as Figure 3.

Prompt:

"Based on this image, generate a visual mockup of the homepage UI for an ecommerce platform selling beauty products and accessories. In the header, replace the Testsite with Dailyessentials with the 'essentials' in red. Replace the hero section's image to a high-quality image of beauty products and accessories. Change 'The Latest Electronics' to 'Move in style this fall' and change 'Get the best deals on the latest tech' to 'Shop the latest trends'. In the Recommended products section, change the Amazing teacup set card to 'Femme Fatale perfume' and replace the image with a perfume bottle. Also change the 'Exquisite Lamp' to 'Cosmic Flare lip gloss' and change the image to a lip gloss bottle. Then the Test cube item, replace that with 'Red Eclipse sunglasses' and change the image to sunglasses. Remove the other items in the Recommended products section. Lastly, Add a footer below the cards and include links for help, about and the social media platforms in the form of icons (specifically Facebook, Instagram and

TikTok). Use the Inter font, rounded buttons, a consistent color palette consisting of white as the primary colour, black as the secondary and red as the accent colour"

Result:

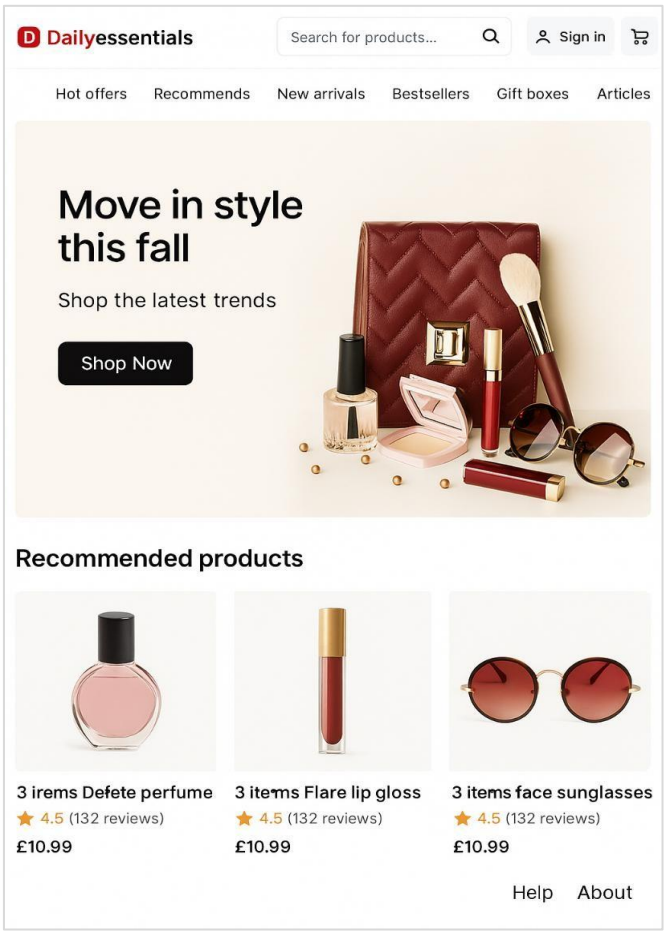


Figure 7: Generated image output from ChatGPT in test 4

2.1.2 Test-5: Markup generation – Direct test

Prompt:

"Generate a markup of a homepage UI for an ecommerce platform selling beauty products and accessories. At the top, include a navigation bar with a logo by the left, category links in the middle, a search bar and shopping cart icon by the right. The hero section should display a large promotional banner with a call-to-action button with text saying 'Shop Now'. Below the hero will be a scrollable section titled 'Featured

products'. It should show 3 items per page and allow the users scroll to the others. The featured items include the femme fatale perfume, the cosmic flare lip gloss, the red eclipse sunglasses, the italian leather handbag and the rose-tinted heels. Each card should include the product image, name and price (£29.99, £5.99, £14.99, £32.57 and

£31.22 respectively). Add a section for trending items in a 4 by 4 grid, populating them with various high-quality images of accessories and beauty products. These items should all have their names and prices below their images, left aligned. The footer should include links for help, about and the social media platforms in the form of icons (specifically Facebook, Instagram and TikTok). Use the Inter font, rounded buttons, a consistent color palette consisting of white as the primary colour, black as the secondary and red as the accent colour"

Result:

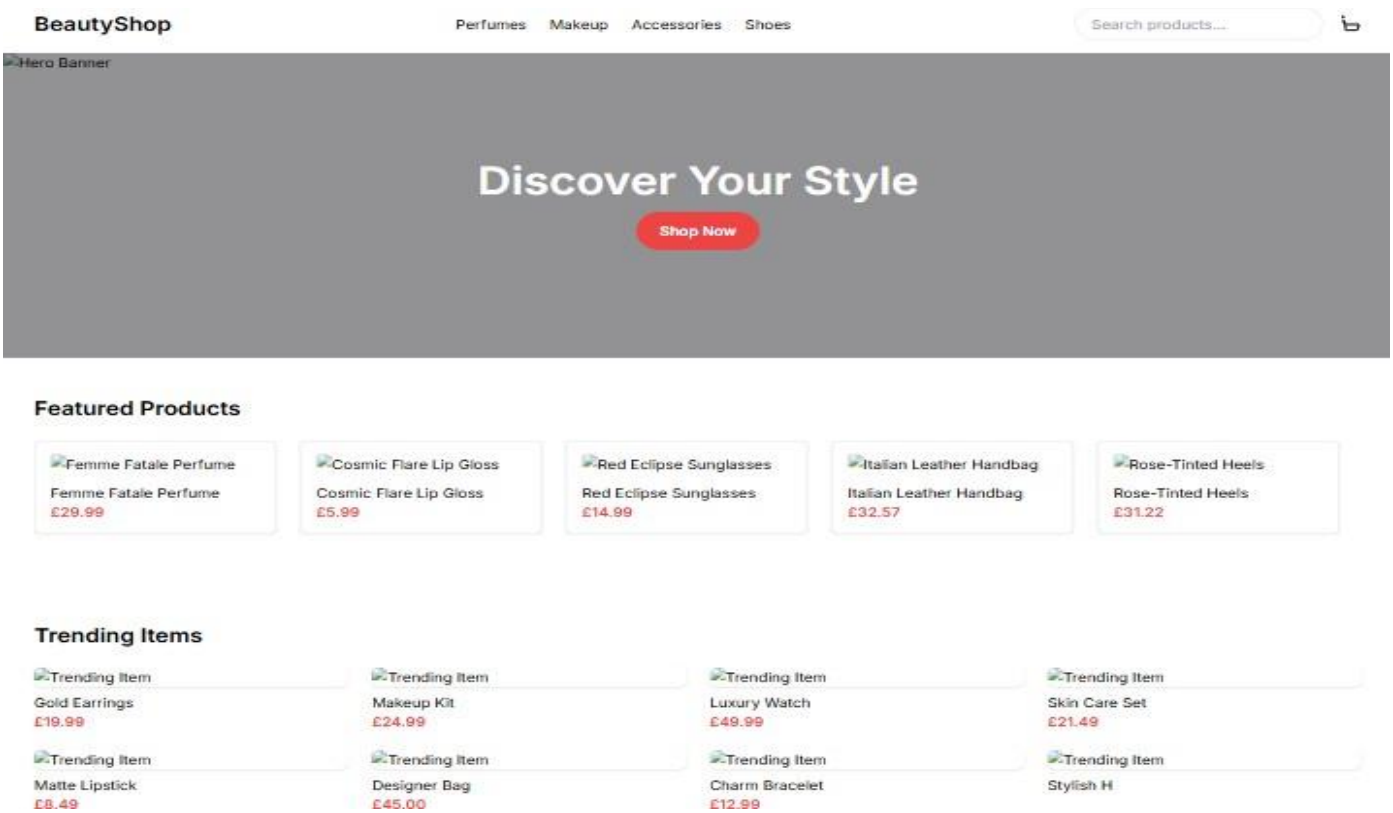


Figure 8: A segment of the markup code render from ChatGPT in test 5

Test-6: Markup generation – Conversion test

Reference image used in this test is shown above as Figure 3.

Prompt:

"Generate the markup and all necessary styling for this user interface"

Result:

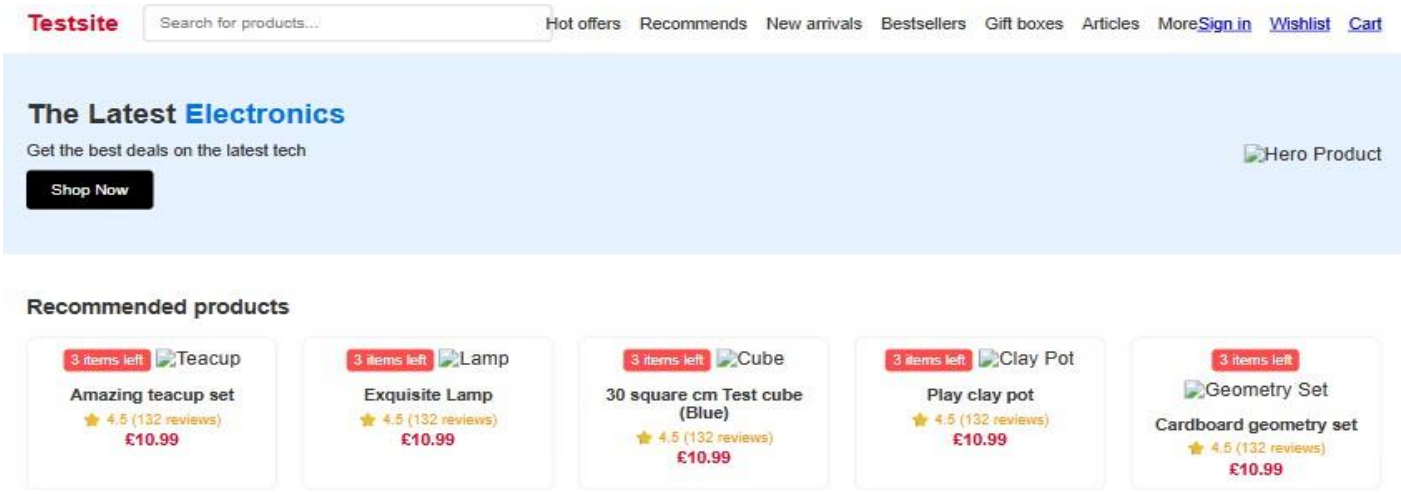


Figure 9: Render of the generated markup from ChatGPT in test 6

Microsoft CoPilot Tests

Test-1: Image generation – Basic prompt

Prompt:

"Generate a visual mockup of the home page user interface for an ecommerce platform"

Result:

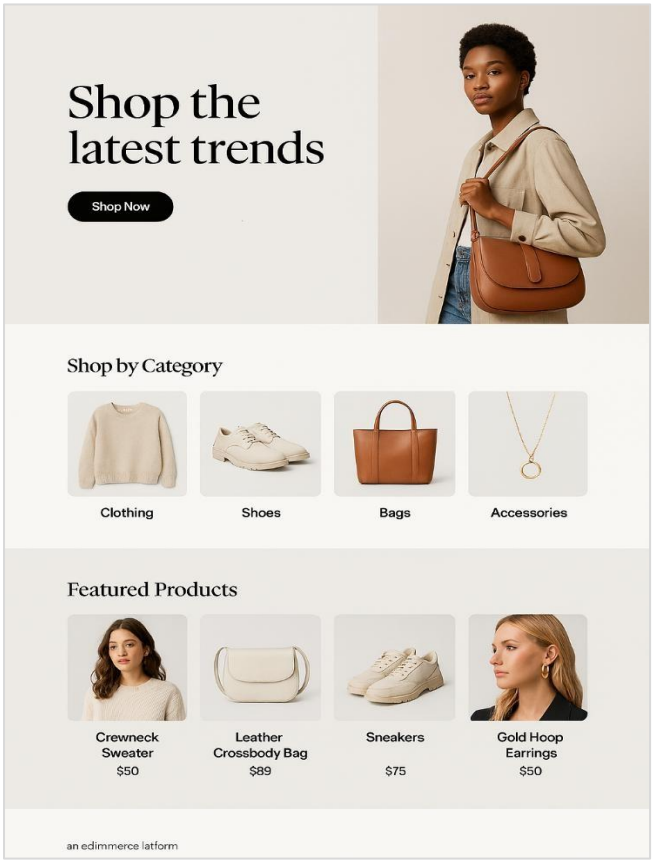


Figure 10: Generated image output from Microsoft CoPilot in test 1

Test-2: Image generation – Intermediate prompt

Prompt:

"Generate a mockup of a home page user interface for an e-commerce platform selling beauty products and accessories. Make the primary colour white and accent red. Showcase various products in this home page but the main featured items are the femme fatale perfume, the cosmic flare lip gloss and the red eclipse sunglasses. Their prices are £29.99, £5.99 and £14.99 respectively."

Result:

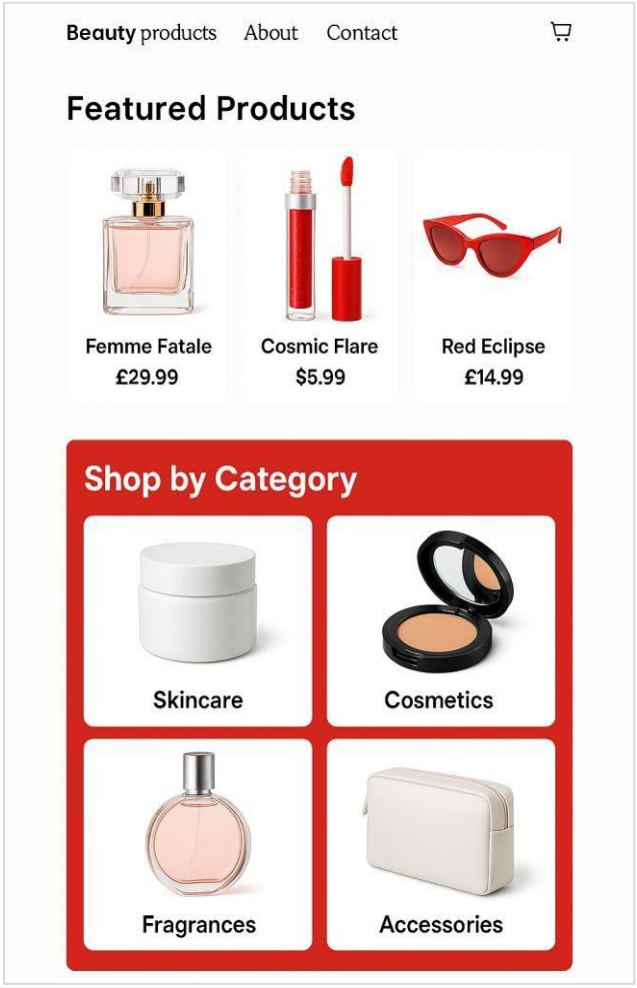


Figure 11: Generated image output from Microsoft CoPilot in test 2

Test-3: Image generation – Advanced prompt

Prompt:

"Generate a visual mockup of a homepage UI for an ecommerce platform selling beauty products and accessories. At the top, include a navigation bar with a logo by the left, category links in the middle, a search bar and shopping cart icon by the right. The hero section should display a large promotional banner with a call-to-action button with text saying 'Shop Now'. Below the hero, showcase featured products which include the femme fatale perfume, the cosmic flare lip gloss and the red eclipse sunglasses, each card including the product image, name and price (£29.99, £5.99 and £14.99 respectively). Add a section for trending items in a 4 by 4 grid, populating them with various high-quality images of accessories and beauty products. These items should all have their names and prices below their images, left aligned. The footer should include links for help, about and the social media platforms in the form of icons (specifically Facebook, Instagram and TikTok). Use the Inter font, rounded buttons, a consistent color palette consisting of white as the primary colour, black as the secondary and red as the accent colour."

Result:

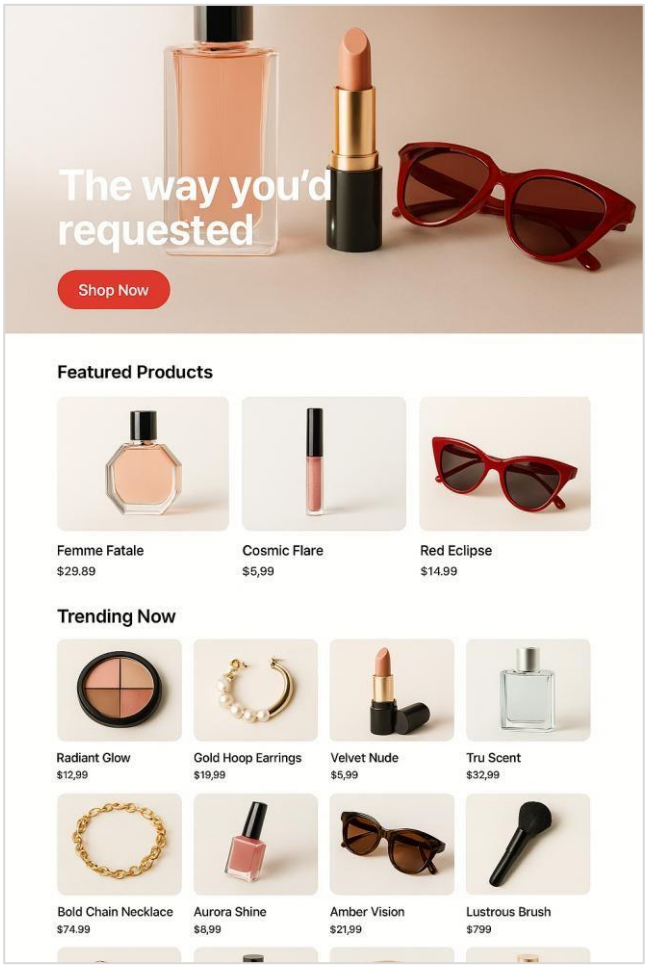


Figure 12: Generated image output from Microsoft CoPilot in test 3

Test-4: Image generation – Transformation test

Reference image used in this test is shown above as Figure 3.

Prompt:

"Based on this image, generate a visual mockup of the homepage UI for an ecommerce platform selling beauty products and accessories. In the header, replace the Testsite with Dailyessentials with the 'essentials' in red. Replace the hero section's image to a high-quality image of beauty products and accessories. Change 'The Latest Electronics' to 'Move in style this fall' and change 'Get the best deals on the latest tech' to 'Shop the latest trends'. In the Recommended products section, change the Amazing teacup set card to 'Femme Fatale perfume' and replace the image with a perfume bottle. Also change the 'Exquisite Lamp' to 'Cosmic Flare lip gloss' and change the image to a lip gloss bottle. Then the Test cube item, replace that with 'Red Eclipse sunglasses' and change the image to sunglasses. Remove the other items in the Recommended products section. Lastly, Add a footer below the cards and include links for help, about and the social media platforms in the form of icons (specifically Facebook, Instagram and TikTok). Use the Inter font, rounded buttons, a

consistent color palette consisting of white as the primary colour, black as the secondary and red as the accent colour"

Result:

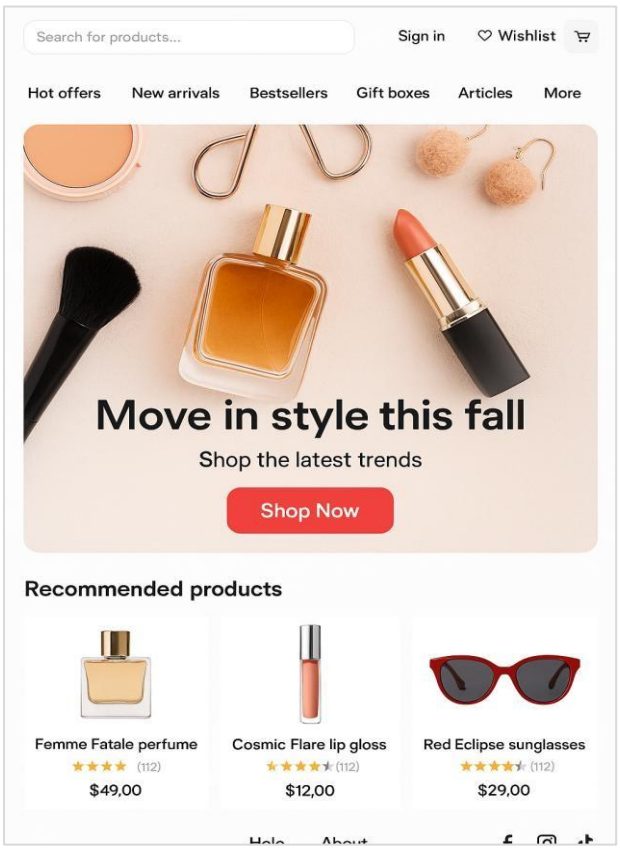


Figure 13: Generated image output from Microsoft CoPilot in test 4

Test-5: Markup generation – Direct test

Prompt:

"Generate a markup of a homepage UI for an ecommerce platform selling beauty products and accessories. At the top, include a navigation bar with a logo by the left, category links in the middle, a search bar and shopping cart icon by the right. The hero section should display a large promotional banner with a call-to-action button with text saying 'Shop Now'. Below the hero will be a scrollable section titled 'Featured products'. It should show 3 items per page and allow the users scroll to the others. The featured items include the femme fatale perfume, the cosmic flare lip gloss, the red eclipse sunglasses, the italian leather handbag and the rose-tinted heels. Each card should include the product image, name and price (£29.99, £5.99, £14.99, £32.57 and £31.22 respectively). Add a section for trending items in a 4 by 4 grid, populating them with various high-quality images of accessories and beauty products. These items should all have their names and prices below their images, left aligned. The footer should include links for help, about and the social media platforms in the form of icons (specifically Facebook, Instagram and TikTok). Use the Inter font, rounded buttons, a consistent color palette consisting of white as the primary colour, black as the secondary and red as the accent colour"

Result:

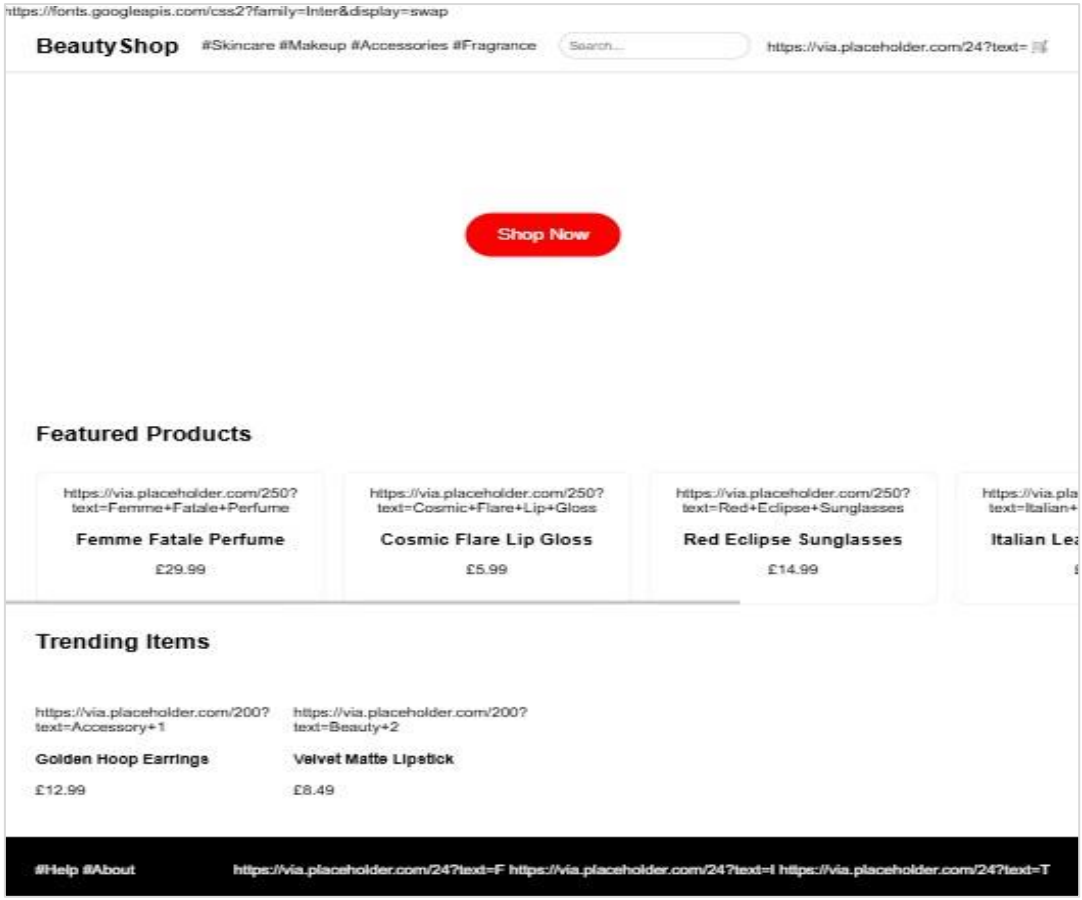


Figure 14: Render of generated markup from CoPilot in test 5

Test-6: Markup generation – Conversion test

Reference image used in this test is shown above as Figure 3.

Prompt:

"Generate the markup and all necessary styling for this user interface"

Result:

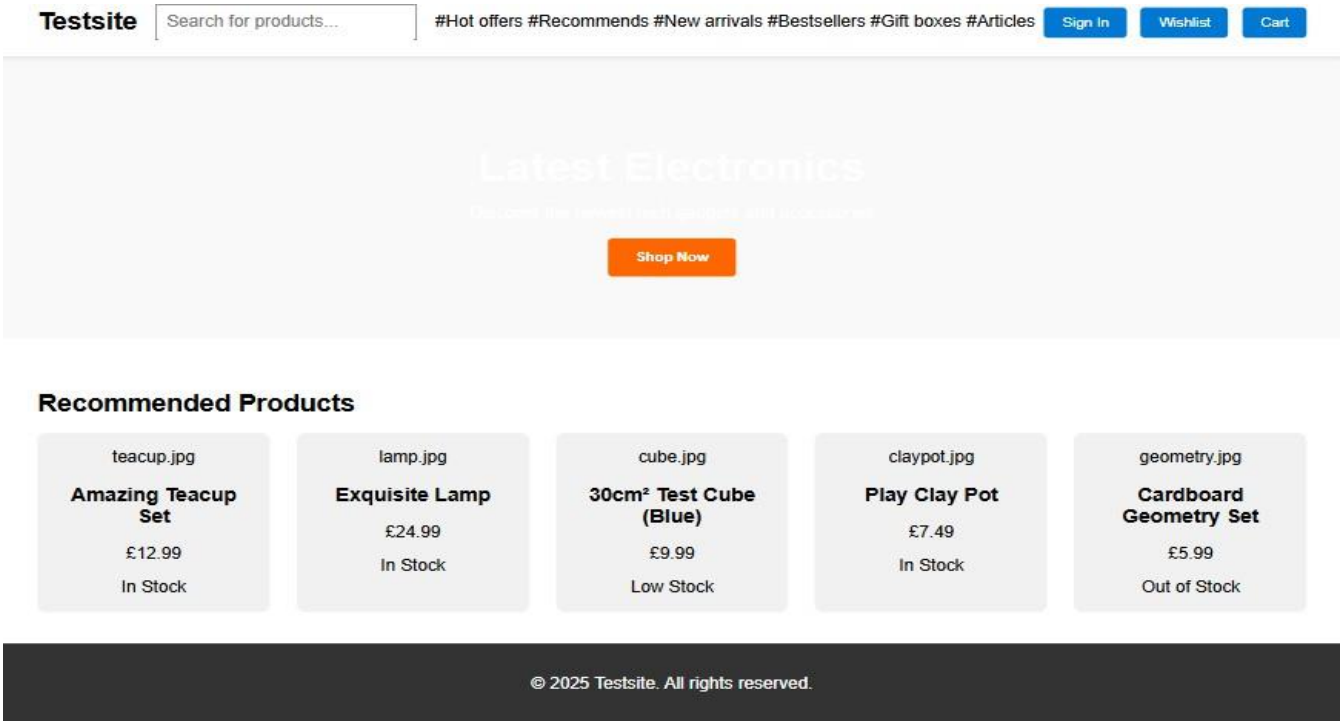


Figure 15: Render of generated markup from CoPilot in test 6

Google Gemini Tests

Test-1: Image generation – Basic prompt

Prompt:

"Generate a visual mockup of the home page user interface for an ecommerce platform"

Result:

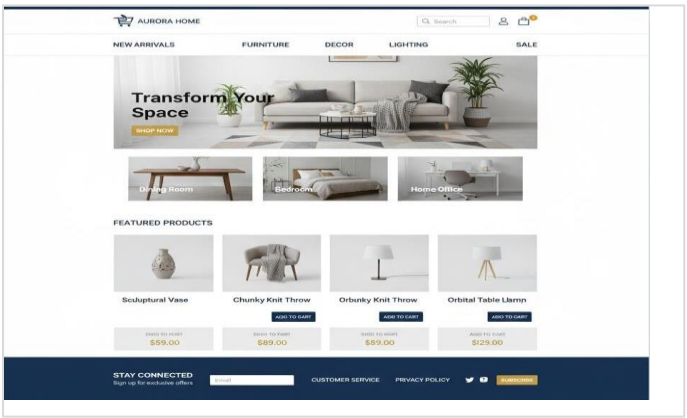


Figure 16: Generated image output from Gemini in test 1

Test-2: Image generation – Intermediate prompt

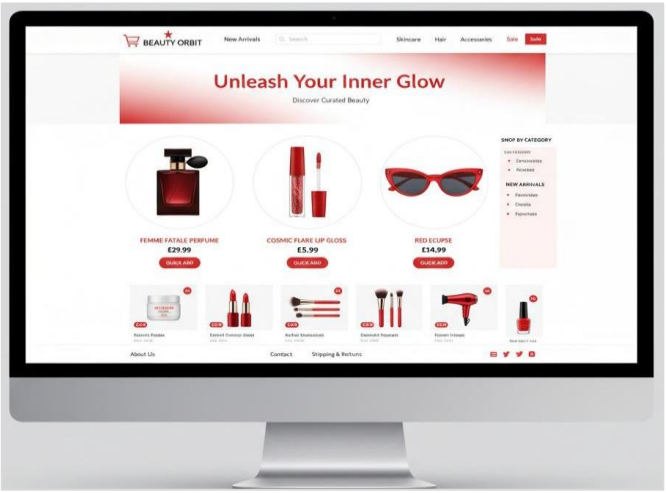
Prompt:

"Generate a mockup of a home page user interface for an e-commerce platform selling beauty products and accessories. Make the primary colour white and accent red. Showcase various products in this home page but the main featured items are the femme fatale perfume, the cosmic flare lip gloss and the red eclipse sunglasses. Their prices are £29.99, £5.99 and £14.99 respectively."

Result:

Figure 17: Generated image output from Gemini in test 2

4.3.1 Test-3: Image generation – Advanced prompt

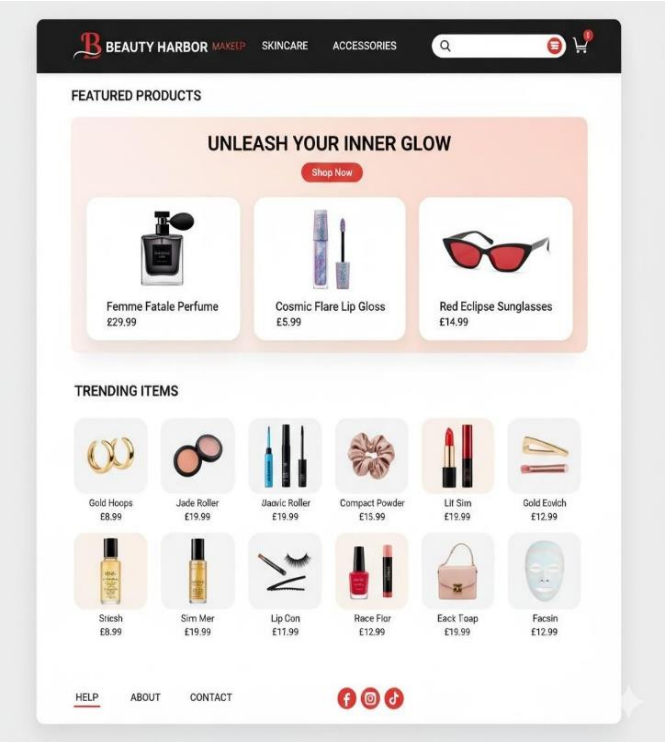


Prompt:

“Generate a visual mockup of a homepage UI for an ecommerce platform selling beauty products and accessories. At the top, include a navigation bar with a logo by the left, category links in the middle, a search bar and shopping cart icon by the right. The hero section should display a large promotional banner with a call-to-action button with text saying ‘Shop Now’. Below the hero, showcase featured products which include the femme fatale perfume, the cosmic flare lip gloss and the red eclipse sunglasses, each card including the product image, name and price (£29.99, £5.99 and £14.99 respectively). Add a section for trending items in a 4 by 4 grid, populating them with various high-quality images of accessories and beauty products. These items should all have their names and prices below their images, left aligned. The footer should include links for help, about and the social media platforms in the form of icons (specifically Facebook, Instagram and TikTok). Use the Inter font, rounded buttons, a consistent color palette consisting of white as the primary colour, black as the secondary and red as the accent colour.”

Result:

Figure 18: Generated image output from Gemini in test 3



Test-3 Analysis

The advanced prompt was designed to test the models' ability to manage a high density of complex and multi layered instructions. The analysis of this test is critical for understanding the extents of GenAI in adhering to detailed design specifications, moving beyond the simpler requirements of the previous tests.

Visual Fidelity:

All models produced aesthetically pleasing outputs. Google Stitch again delivered the most polished design again, with a premium aesthetic that perfectly matched the high-end beauty theme. Microsoft CoPilot produced a visually striking mockup with a high-quality, photographic hero image, but its aesthetic quality is undermined by its failure to follow basic structural prompts. Google Gemini generated a clean, modern, and well-organised design with consistent use of rounded corners as requested. ChatGPT’s output remained consistent with its minimalist style, which, while clean, felt less complete due to the omitted footer.

Layout and Structural Coherence:

The structural coherence of the outputs was directly tied to prompt adherence. Google Stitch and Google Gemini produced the most coherent and complete page layouts, as they included all the fundamental components of a webpage from header to footer. Their structures were logical and followed professional design patterns. The layouts from ChatGPT and Microsoft CoPilot, however, must be considered structurally incomplete. By omitting the footer (and in CoPilot's case, the header), their outputs represent fragments of a webpage rather than a coherent whole, a significant flaw that would require major developer intervention to rectify.

Prompt Adherence:

This test revealed significant differences in the models' ability to follow complex instructions. Google Stitch and Google Gemini demonstrated the strongest adherence, successfully incorporating most of the key structural components, including the detailed navigation bar and, crucially, the footer with social media icons. In contrast, ChatGPT and Microsoft CoPilot both failed to generate the requested footer, a major omission that points to a limitation in processing a long series of commands. Furthermore, CoPilot also omitted the entire navigation bar, a fundamental failure of adherence. Notably, a universal failure point was the request for a "4 by 4 grid" for trending items; all models produced a two-row grid, suggesting a common difficulty in interpreting specific, complex spatial layout commands.

Summary:

GenAI Tool	Visual Fidelity	Layout Coherence	Prompt Adherence	Remarks
ChatGPT	3	4	3	Clean design, but failed to adhere to key instructions like the footer and grid size.
Microsoft CoPilot	2	4	2	Aesthetically pleasing but failed on fundamental instructions, omitting both header and footer.
Google Gemini	4	4	5	Strong adherence and a complete page structure, though it misinterpreted the grid layout.
Google Stitch	4	5	5	Excellent, professional design that followed nearly all instructions, only failing on the grid size.

Test-4 Analysis

This test assessed the advanced, multimodal capabilities of the GenAI models by requiring them to interpret a source UI image and apply a series of complex textual modifications. The primary goal was to evaluate their ability not just to generate from scratch, but to understand, deconstruct, and selectively edit an existing design, a task that more closely mirrors a real-world redesign workflow.

Visual Fidelity:

All models were successful in producing high-quality, aesthetically pleasing images that matched the target theme of a beauty e-commerce site. The generated hero images and product shots were professional across the board. Google Stitch's output was again the most polished, presenting a cohesive and professionally executed redesign. The outputs from the other models were also visually strong; however, this aesthetic quality often masked significant underlying failures. For instance, while ChatGPT's result was visually clean, but it contained squeezed tests.

Layout and Structural Coherence:

The structural coherence of the final designs was entirely dependent on each model's success in executing the transformation. As Google Stitch performed the edits correctly, its final layout was logical and coherent. Microsoft CoPilot's layout was also largely coherent, but the jarring inconsistency of the unchanged header from the original "Testsite" theme undermined the integrity of the redesign. The layouts from ChatGPT and Google Gemini were the least coherent due to their failure to correctly replace and remove elements, resulting in a cluttered and illogical presentation of products that would potentially confuse a user.

Prompt Adherence:

The transformation task proved to be the most challenging test yet, revealing significant variations in multimodal understanding. Google Stitch demonstrated the highest level of adherence, successfully executing nearly all transformations: it changed the brand name, replaced the hero section, correctly substituted the three specified products,

removed the extraneous items, and added the complete footer. In stark contrast, ChatGPT and Google Gemini both failed on the critical instruction to remove the original products, resulting in a confused mixture of old and new elements. Furthermore, ChatGPT generated garbled text for the product names (e.g., "3 irems Defete perfume") and only partially added the footer, omitting the required social media icons. Microsoft CoPilot performed moderately, successfully replacing and removing products and adding the footer, but it completely failed to alter the header as instructed. A notable finding was the universal failure of all four models to adhere to the fine-grained instruction to colour the word "essentials" red in the logo, suggesting a common limitation in detailed, in-image text manipulation.

Summary:

GenAI Tool	Visual Fidelity	Layout Coherence	Prompt Adherence	Remarks
ChatGPT	2	4	2	Visually clean but failed on multiple key edits and generated nonsensical text.
Microsoft CoPilot	3	4	4	Succeeded in replacing content but completely failed to edit the header as instructed.
Google Gemini	2	4	2	Failed to correctly replace and remove elements, resulting in a confused and incoherent layout.
Google Stitch	5	5	5	Near-perfect execution of the complex transformation task,

Test-5 Analysis

This test evaluated the models' capability to generate a complete and functional UI from a detailed textual description, a critical task for integrating GenAI into a development workflow. The prompt was identical in complexity to the advanced image prompt but required the output to be HTML and CSS markup.

Visual Fidelity:

The visual quality of the rendered UIs varied dramatically. The output from Google Stitch was the only one that met professional design standards, presenting a clean, modern, and aesthetically pleasing interface ready for production. The render from ChatGPT was rudimentary; while functional, it lacked any sophisticated styling beyond the basic defaults provided by its framework. The results from CoPilot and Gemini cannot be meaningfully assessed on aesthetics due to their catastrophic functional and structural failures, with CoPilot's render being visibly broken and Gemini's being completely irrelevant to the prompt.

Code Quality:

A direct review of the code highlights paradoxical results. Google Stitch produced the highest quality code, using the Tailwind CSS framework (a professional framework in web development) on top of a fully semantic HTML structure like <header>, <nav>, and <main>. Surprisingly, Google Gemini, despite failing the prompt, produced exceptionally high-quality code for what it did generate, using a clean semantic structure and well-organized, responsive CSS with variables. ChatGPT's code was decent, also using Tailwind and a mostly semantic structure, but it was functionally incomplete. The code from Microsoft CoPilot was of extremely low quality, consisting of malformed HTML with un-rendered tags and poorly structured CSS, making it completely unusable for a developer.

Accessibility Guidelines Alignment:

The models' understanding of web accessibility also showed a wide range of sophistication. Google Stitch and Google Gemini both demonstrated a strong grasp of accessibility principles, correctly implementing aria-label attributes on interactive icons to make them understandable to screen readers. ChatGPT, however, failed on accessibility basics, critically omitting alt attributes from all of its <img> tags. Microsoft CoPilot's code was too broken to be evaluated for accessibility, as it would be fundamentally inaccessible to all users.

Prompt Adherence:

Evaluating the results revealed significant performance gaps. Google Stitch came the closest to executing the prompt, correctly implementing most structural components, including the header, hero, a fully functional scrollable "Featured Products" section, and a detailed footer. However, it failed on the specific instruction for the "Trending Items" section, producing a 2x4 grid instead of the requested 4x4 layout. ChatGPT produced a structurally plausible page but failed on two key requirements: the featured products section was not scrollable, and it only generated 8 of the 16 requested trending items. Google Gemini fundamentally misinterpreted the prompt, ignoring the complex instructions entirely. The output from Microsoft CoPilot was a total failure, producing broken markup that did not adhere to any of the prompt's structural or functional requirements. Notably, this test showed that no model could correctly interpret the "4 by 4 grid" instruction.

Summary:

GenAI Tool	Visual Fidelity	Code Quality	Accessibility Alignment	Prompt Adherence	Remarks
ChatGPT	3	4	2	3	Produced a basic structure but failed on key functional requirements and lacked styling.
Microsoft CoPilot	1	1	1	1	Total failure; the generated markup was broken and resulted in an unusable UI.
Google Gemini	2	4	4	1	Completely misinterpreted the prompt, generating a non-functional page with incorrect elements.
Google Stitch	5	5	4	4	Exceptional performance and codebase, but failed on the specific grid layout instruction.

Summary of Study

This dissertation investigated the feasibility of Generative AI to generate quality user interfaces. A dual-method experiment across four AI tools was conducted. Amongst these four tools, three are general multipurpose AI models and one (Google Stitch) is specialised for building UIs. The study's findings provide some clear and nuanced answers to this question.

The evidence points to three primary conclusions. First, a significant "ideation-implementation gap" exists: generalist AI tools are more effective at visual brainstorming, producing aesthetic images as UI designs. On the other hand, they struggle at generating the functional, developer-ready code to match.

Secondly, model specialisation appears critical. The generalist models (ChatGPT, CoPilot, Gemini) consistently failed the markup generation tests, while the purpose-built UI generator (Google Stitch) was the only tool capable of reliably producing high-fidelity and accessible code from the prompts.

Finally, the study identified a "complexity ceiling" for the models, which begin to omit specifically requested items as prompts became more detailed. This, combined with the finding that prompt engineering is a non-linear skill (where partial detail can yield worse results than simple prompts), suggests that precise control over generalist AI is not fully reliable currently.

CONCLUSION AND RECOMMENDATIONS

Generative AI is not a single entity but a spectrum of tools with vastly different capabilities. First, for designers, generalist AI is already a powerful assistant for visual ideation. Second, and more critically, for developers, these same tools are largely unreliable as far as implementing generated UIs go.

The true, tangible future of AI-driven development is not in the "do-it-all" models, but in the highly specialised, domain-trained model. The evidence from this research shows that such tools are already capable of translating design into high-quality code. The question is no longer "if" GenAI tools can generate usable UI, but rather how quickly the industry will adopt the specialised tools that are already making it a reality.

REFERENCES

1. Aktar, Z. B., 2024. Unveiling the evolution of generative AI (GAI): a comprehensive and investigative analysis toward LLM models (2021–2024) and beyond. Journal of Electrical Systems and Information Technology, 11(22).
2. Goodfellow, I. et al., 2014. Generative Adversarial Nets. Montreal, QC, NeurIPS Foundation.
3. Lawton, G., 2025. What is GenAI? Generative AI explained | TechTarget. [Online] Available at: https://www.techtarget.com/searchenterpriseai/definition/generative-AI [Accessed September 2025].
4. Marr, B., 2024a. How Generative AI Will Change

The Jobs Of Artists And Designers. [Online] Available at: https://www.forbes.com/sites/bernardmarr/2024/05/28/how-generative-ai-will-change-the-jobs-of-artists-and-designers/ [Accessed 3 October 2025].

5. Marr, B., 2024b. The 4 Types Of Generative AI Transforming Our World. [Online] Available at: https://www.forbes.com/sites/bernardmarr/2024/04/29/the-4-types-of-generative-ai-transforming-our-world [Accessed 30 September 2025].

6. Moran, K. & Gibbons, S., 2024. Generative UI and Outcome-Oriented Design. [Online] Available at: https://www.nngroup.com/articles/generative-ui/ [Accessed 12 August 2025].

7. Nielsen, J., 2024. 10 Usability Heuristics for User Interface Design. [Online] Available at: https://www.nngroup.com/articles/ten-usability-heuristics/ [Accessed 2 October 2025].

8. Norman, D., 2013. The Design of Everyday Things: Revised and Expanded Edition. s.l.:Basic Books.

9. Okopnyi, P., Nordberg, O. E. & Guribye, F., 2024. Against Generative UI. s.l., HttF '24, pp. 1-4.

10. Shneiderman, B., 2020. Human-Centered AI: A framework for reliable, safe, and trustworthy systems. International Journal of Human–Computer Interaction, pp. 495-504.

11. Wan Husain, W. N. L. & Jantan, A. H., 2017. Adaptive User Interface Design: A Case Study of Web Recommendation System. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 20 October, 9(3-3), pp. 137-140.