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)

# GenAI Tools and Coherence of Generated User Interface Design

**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). *GenAI tools and coherence of generated user interface design*. *UKR Journal of Engineering, Innovation and Technology, 1*(1), 1-7.

**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 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 quality, usability, and coherence of generated interfaces that are not only functional but also meet usability and accessibility standards.

### Objectives

The study is guided by the following objectives:

1. To evaluate the quality, usability, and coherence of generated interfaces, using known industry standards and guidelines that govern UI design.

## LITERATURE REVIEW

### Understanding User Interfaces and Generative AI

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.

**Current Applications of AI in UI Design**

**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.

**Future Directions: Generative UI**

The emergence of Generative User Interfaces (GenUI) marks a shift on how digital interfaces are designed and experienced. Unlike AI-assisted design, which enhances the designer's workflow, or AI-personalised interfaces, which adapt pre-existing layouts, Generative UI seeks to dynamically modify interfaces in real time based on user intent and goals (Moran & Gibbons, 2024; Okopnyi, et al., 2024). GenUI systems move away from static designs toward outcome-oriented designs as interfaces are generated to help users achieve specific objectives on the fly. Designers define guidelines and constraints like accessibility aspects, brand identity, etc., then the AI systems generate UI components that best satisfy the user's immediate needs.

Generative UI builds upon advances in large language models (LLMs), enabling systems to understand natural language inputs and translate them into usable interface layouts. For instance, a generative interface could construct a custom analytics dashboard when a user requests, "show trends for sales across regions this quarter," automatically generating charts, filters, and controls relevant to the task. However, Moran and Gibbons (2024) point out several challenges accompany this transition, including consistency across sessions, the cognitive cost of changing layouts, and the risk of reduced user trust if the system's generative decisions lack transparency. Okopnyi, Nordberg and Guribye (2024) also note in their paper that by removing a human designer from the process, this skips the evaluation of generated UIs, relying solely on a user's feedback.

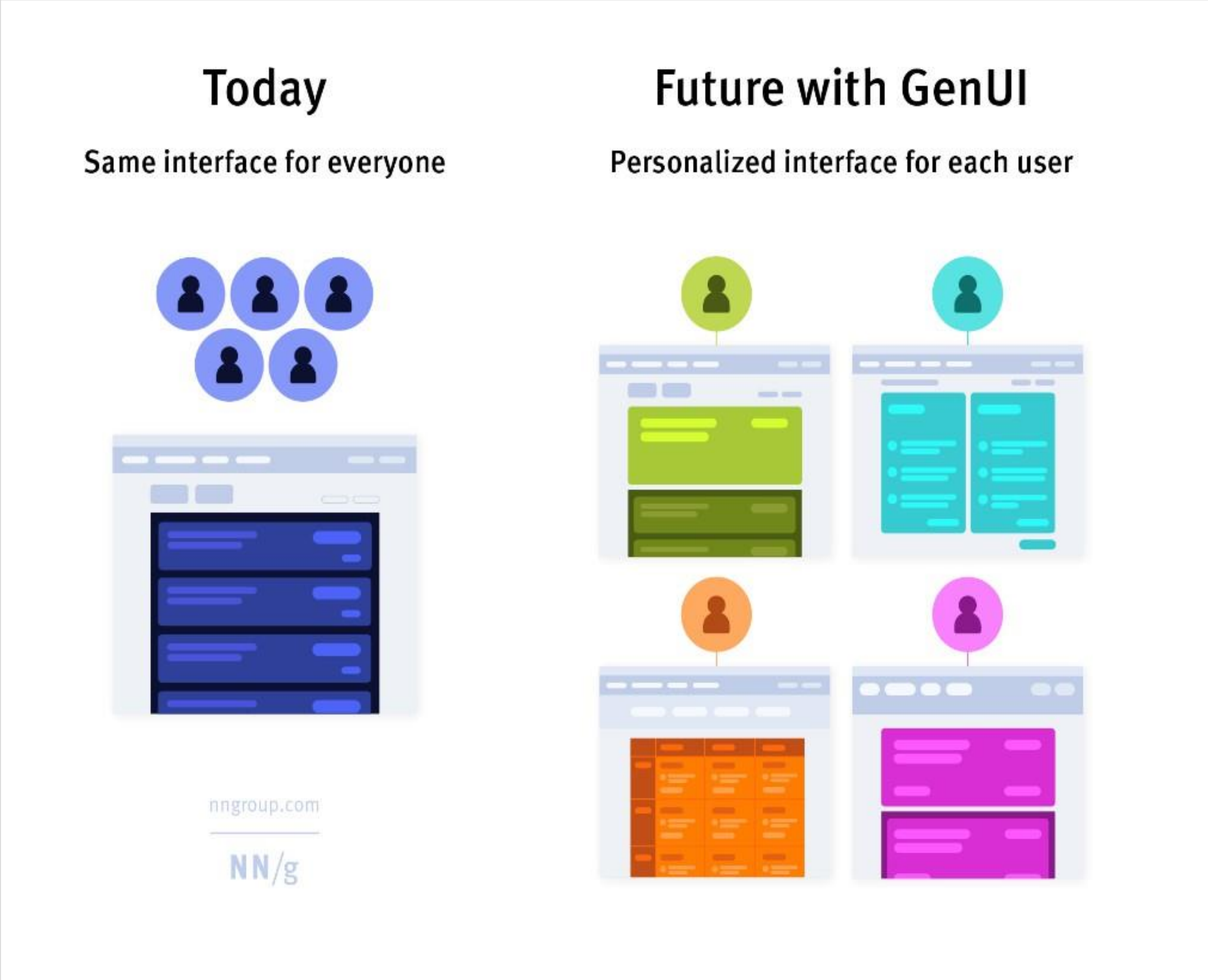


*Figure 2: An illustration of GenUI adapting different layouts to fit users' needs (Moran & Gibbons, 2024)*

In essence, Generative UI represents the convergence of automation and personalisation. A future in which interfaces become fluid and responsive entities rather than fixed constructs as illustrated in **Figure 2**. As Shneiderman (2020) emphasises, such a shift must remain grounded in human-centred AI principles to ensure that usability, accessibility, and user agency remain central as interface generation becomes more autonomous. However, Okopnyi, Nordberg and Guribye (2024) suggest this may be a shift too far into the human-centred principles with no guarantee it will suite every type of user.

## 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

**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

## IMPLEMENTATION

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:*

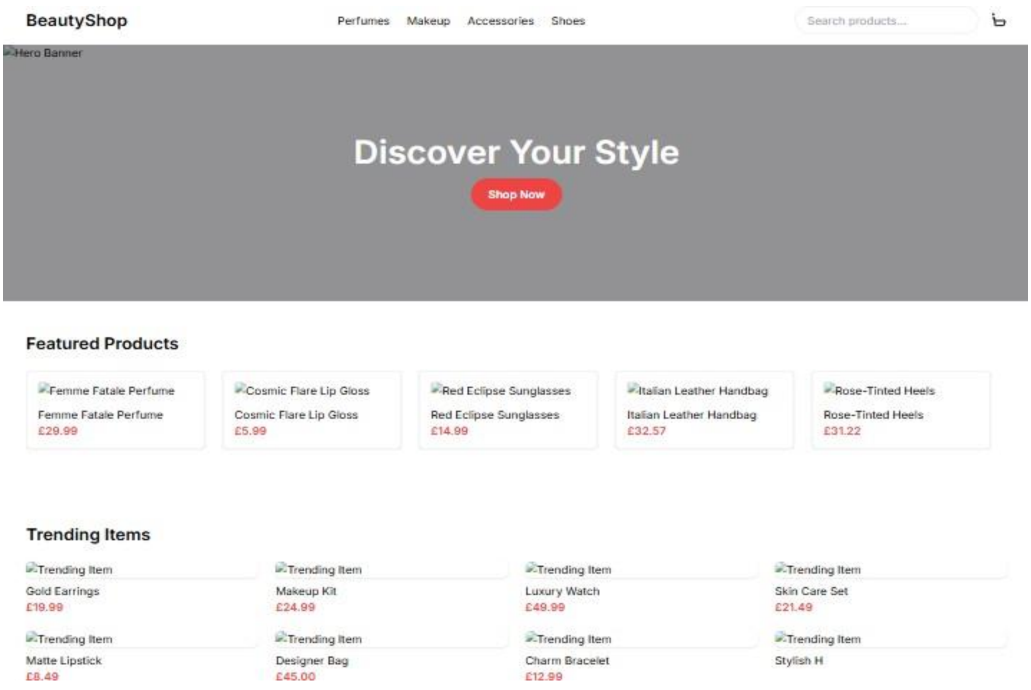


*Figure 8: A segment of the markup code render from ChatGPT in test 5*

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:*

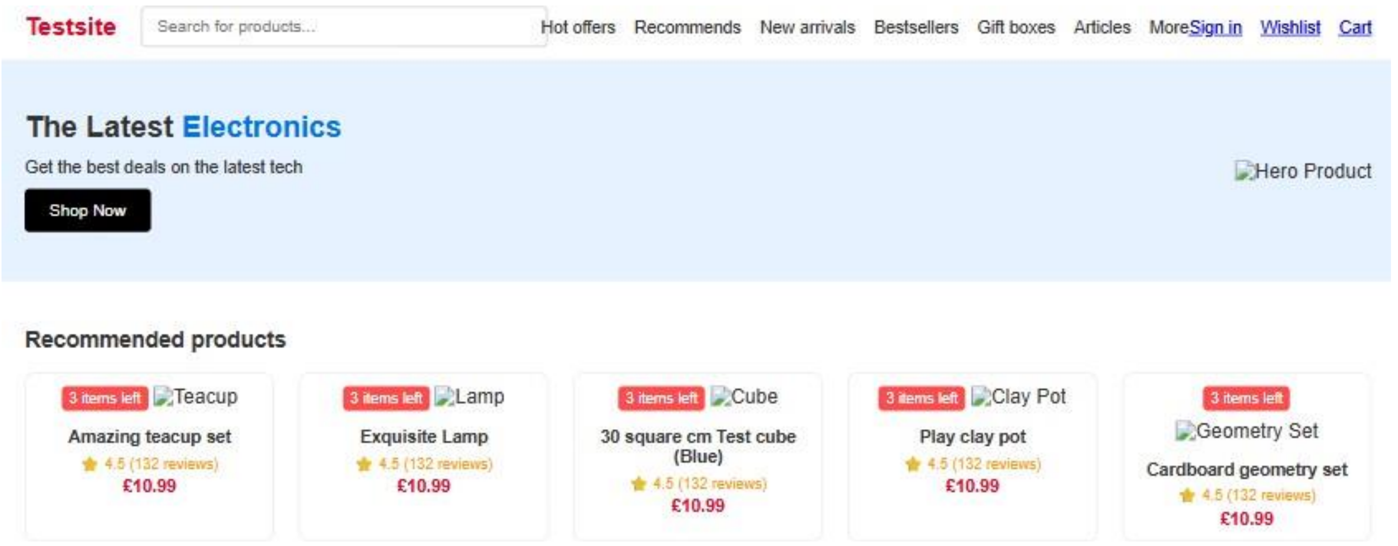


*Figure 9: Render of the generated markup from ChatGPT in test 6*

**Test-2 Analysis**

This test upped the complexity from the baseline, introducing specific content, colour, and pricing elements into the prompts. The analysis evaluates how effectively each model adhered to these intermediate-level design prompt which helps to understand the kind of results from less seasoned designers in prompting GenAI.

*Visual Fidelity:*

The aesthetic quality of the outputs showed significant divergence. Google Stitch produced the most polished and professional result, with a high-end feel appropriate for

a premium beauty brand. Its use of colour and elegant typography effectively builds a sense of brand quality, although the hero section's graphic doesn't seem to fit with the style of the other generated images, making it to stick out. ChatGPT's output was clean and minimalist, with a good use of whitespace that made the featured products the clear focus. Microsoft CoPilot's design was bold with a large solid red block enveloping the categories section. This could be considered visually overpowering and less refined compared to the other results. The design from Google Gemini, while featuring all the required elements, appeared the most cluttered and dated, with an unconventional layout and gradients that detract from a modern aesthetic.

*Layout and Structural Coherence:*

In terms of structure, all models produced logically coherent layouts. Google Stitch and Google Gemini generated the most comprehensive page structures, including headers, hero sections, multiple product sections, and footers, making them feel like more complete webpages. The layouts from ChatGPT and Microsoft CoPilot were more minimal, focusing primarily on the requested "Featured Products" and "Shop by Category" sections without including standard navigational elements like a header or footer. While these simpler layouts were still easy to understand, their incompleteness would make them less viable as a direct starting point for a full homepage design.

*Prompt Adherence:*

All four models demonstrated a strong capability to adhere to the specific content constraints of the intermediate prompt. Each successfully generated a UI featuring the "femme fatale perfume," "cosmic flare lip gloss," and "red eclipse sunglasses," including their respective prices. The interpretation of the "accent red" colour instruction varied, with ChatGPT incorporating it subtly through product imagery, while Microsoft CoPilot used it boldly in a large background block. Both Google

Gemini and Google Stitch also integrated the red accent effectively, demonstrating that all models could translate explicit stylistic instructions into their outputs.

*Summary:*

| GenAI Tool | Visual Fidelity | Layout Coherence | Prompt Adherence | Remarks |
|---|---|---|---|---|
| ChatGPT | 5 | 4 | 5 | Clean, minimalist design that adheres to the prompt but lacks a complete page structure. |
| Microsoft CoPilot | 4 | 3 | 5 | Follows prompt with a bold design, but the aesthetic is less refined and the layout |

| | | | | |
|---|---|---|---|---|
| | | | | is incomplete. |
| Google Gemini | 3 | 3 | 5 | Adheres to all constraints with a comprehensive layout, but the visual design appears cluttered and dated. |
| Google Stitch | 5 | 5 | 5 | Excellent execution on all criteria, producing a polished, professional, and brand-aligned design. |

## DISCUSSION

The implementation and evaluation chapters provide empirical data on the current capabilities of Generative AI in UI generation. The results from the six tests, performed across four different models, reveal some insights into the relationships between the prompts, the AI tools and the practicality of the results. The points discussed below are structured around the four primary patterns that emerged from the data: the limitations of prompt complexity, the gap between visual ideation and functional implementation, the critical role of model specialisation, and the non-linear nature of prompt engineering.

### The Complexity Ceiling: A Correlation between Detail and Unreliability

A consistent finding across the generalist models (ChatGPT, Microsoft CoPilot, and Google Gemini) was the degradation of performance as prompt complexity increased. While all models successfully interpreted the simple and intermediate prompts of Test 1 and Test 2, the advanced prompt of Test 3 proved to be a significant challenge.

The repeated failure of both ChatGPT and Microsoft CoPilot to generate the requested footer in Test 3 is a critical data point. This omission suggests a "ceiling on complexity," or a functional limit to the number of discrete constraints these models can reliably track and execute in a single generation. This limitation was further exemplified by the universal failure of all models to correctly interpret the "4 by 4 grid" instruction, defaulting to simpler, more common 2-row layouts.

This implies that these tools in their current state are effective at executing simple requests. Simpler requests may allow the tools to make more familiar assumptions and thus produce generations closer to what they were trained on. More contextual prompts may produce more specific outputs but are not yet fully reliable since they tend to miss out on some details.

### 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.