International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

# AI Enabled Virtual Try-On for E-Commerce

**B. Dhanu Raj, G. Manasa, A. Chandra Vamsi, D. Uday Kiran**
*22BQ1A5420 22BQ1A5441 22BQ1A5408 22BQ1A5436*
**Mr. V. Gopinath**, Assistant Professor
*Department of Artificial Intelligence & Data Science*
*Vasireddy Venkatadri Institute of Technology (Autonomous)*
*Affiliated to JNTUK | NAAC 'A' Grade | Guntur, Andhra Pradesh - 522 508*

***Abstract*** *— In today's digital environment, online fashion sales face return rates of 30 to 40 percent because customers cannot try products or things before buying. This paper introduces StyleGenie, an AI-enabled virtual try-on system that allows users to see photorealistic product images on their own photos The software uses Google Gemini generative multimodal AI model architecture for generating realistic images in more than nine categories of products including eyewear, shirts, hoodies, t-shirts, watches, women's tops, and traditional kurtas. A multi-model fallback strategy achieves an average try-on generation success rate of 92.9 percent. The platform offers AI-driven style analysis, which includes face shape detection, fit scoring, color harmony evaluation, and a stylist's recommendation. Built on React 18 with TypeScript and a secure Node.js/Express backend proxy, the system prioritizes user privacy and does not store images permanently. Client-side image processing reduces file size by over 70 percent. A pilot evaluation with 35 participants results in a System Usability Scale (SUS) score of 84.2, which is considered 'Excellent,' along with a 76 percent self-reported decrease in purchase uncertainty.*

*Index Terms — Virtual Try-On, Generative AI, E-Commerce, Google Gemini, Fashion Technology, Computer Vision, React.js, Image Synthesis, Personalization, Style Analysis.*

## I. INTRODUCTION

The evolution of electronic commerce has significantly changed the retail industry Online shopping has experienced rapid growth in the fashion and apparel industry, however, unlike the physical store, it does not give an opportunity for users to physically try out the product before making a purchase. Consequently, there is uncertainty regarding appearance, fit, and suitability of items leading to returns of 30 to 40 percent in the international online fashion market.

Virtual Try-On (VTO) technology has been innovatively created. A VTO utilizes Artificial intelligence,computer vision and generativeComputer Vision to show customers the appearance of the item in an online environment. Through analyzing the user's face/body features, such systems provide realistic overlays of clothing or accessories onto the user's image allowing users to feel comfortable making purchase decisions.

The latest advancements in options. The introduction of generative AI using multi-modal transformers such as Google Gemini has enhanced this simulation by providing highly realistic images with the maintenance of proper alignment, lighting and structure. These transformer models are able to use user images and product images alongside text instructions for optimal performance.

The proposed system, StyleGenie, AI-Enabled Virtual Try-On for E-Commerce, will present an integrated online platform. Users can explore fashion items, submit or upload their photograph, experience realistic try-ons generated by AI, get a comprehensive style analysis and easily change between before and after pictures with an immediate ability to download vision,options.artificial intelligence.

## II. LITERATURE REVIEW

### *A. Image-Based Virtual Try-On Approaches*

Han et al. [2] introduced VITON, the first image-based virtual try-on network that uses CNNs to match clothing with user body structure. Wang et al. [5] expanded on this with CP-VTON, which maintains garment texture through geometric matching. Choi et al. [6] developed VITON-HD, achieving higher-resolution outputs with misalignment-aware normalization. However, all these methods are costly in terms of computation, need large paired training datasets, and are limited to upper-body clothing To solve these problems, StyleGenie uses the Gemini AI model on cloud-based generative AI that does not need any special training data sets and supports nine diverse product categories.

### B. *Augmented Reality Try-On Systems*

Commercial augmented reality (AR) systems, including Lenskart's eyewear try-on and Snap Inc.'s AR lenses, utilize facial landmark detection and 3D mesh overlays to enable real-time visualization [7]. Although these systems deliver rapid and interactive experiences, their application is restricted to specific product categories, primarily eyewear and cosmetics, and they do not offer styling recommendations. Karthik et al. [8] introduced a computer vision-based virtual dressing room employing image overlay, which enhanced user interaction but did not achieve realistic results under varying lighting conditions.

### C. *Generative AI for Visual Synthesis*

Sharma et al [9] demonstrated a VTO system based on AI, which leveraged deep learning technology to generate try-on images, resulting in greater personalization than typical systems but requiring more computational power locally. Recent improvements, such as Google Gemini's multimodal transformer abilities [1], Stable Diffusion-based Latent Diffusion Models [11], and DALL-E 3 [10], have made image synthesis that follows instructions much more realistic. These models can take multi-image prompts with text instructions and keep the user's identity while accurately showing the product's features.

### D. *Privacy in AI-Powered Applications*

Privacy-preserving AI research stresses that applications that process biometric data must follow three main rules: they must minimize data, process it on the device, and not keep it at all [12]. Most current VTO systems upload and keep user photos for processing, which is a significant privacy concern. StyleGenie solves this problem by using a stateless architecture, which means that images only exist in the browser's temporary memory during the active processing session and are then permanently deleted.

### E. *Gap Analysis*

No previous work has combined generative AI, multi-category product support, and AI-driven style analysis with client-side privacy preprocessing and a production-ready web deployment into a single system that non-technical users can use. StyleGenie is the first system made by students that can do this full-stack integration while still being free to use in the cloud.

## III. PROPOSED SYSTEM

### A. *System Overview*

StyleGenie is a full-stack web application structured around three main subsystems: the AI Try-On Engine, which creates photorealistic images; the AI Style Analysis Module, which gives personalized fashion feedback; and the Privacy-First Processing Pipeline, which handles client-side images without storing them permanently.

**TABLE I**

*Comparison of StyleGenie with Existing Virtual Try-On Systems*

| System | Categories | Real-Time | Custom Upload | AI Styling | Privacy |
|---|---|---|---|---|---|
| VITON (Han,2018) | Upper-body | No | No | No | No |
| CP-VTON (Wang,2018) | Upper-body | No | No | No | No |
| VITON-HD (Choi,2021) | Clothing only | No | No | No | No |
| Lenskart AR | Eyewear only | Yes | No | No | No |
| Snap Inc. AR | Cosmetics/ EW | Yes | No | No | No |
| Sharma et al.(2021) | Limited | No | No | No | No |
| StyleGenie (Proposed) | 9+ Categories | Yes | Yes | Yes | Yes |

### B. *Key Features*

- Multi-product category inventory: spectacles, shirts, hoodies, tees, watches, tops for women, and kurtas.
- Selfie webcam, with face alignment prompts, 3-second countdown, and horizontal alignment prompts.
- Custom product addition not limited by the above-mentioned inventory list.
- Photographically realistic try-on with Google Gemini multimodal generative model.
- Multi-generative model support: gemini-2.5-flash-image → gemini-3.1-flash-image-preview with 3 attempts in case of failure.
- STYLE analysis with AI: face shape identification, fit rating, color coordination, and Stylist's Verdict.
- Face gender recognition to improve recommendations.
- Privacy-first: no user images stored on any server at any time.

## IV. METHODOLOGY

### *A. System Architecture*

The system has three levels: (1) the Client Layer, which is a React SPA that handles webcam streaming and client-side preprocessing; (2) the Proxy Layer, which is a Node.js/Express server that securely injects the API key and enforces rate limiting; and (3) the AI Service Layer, which is the Google Gemini cloud API that performs multimodal image synthesis.

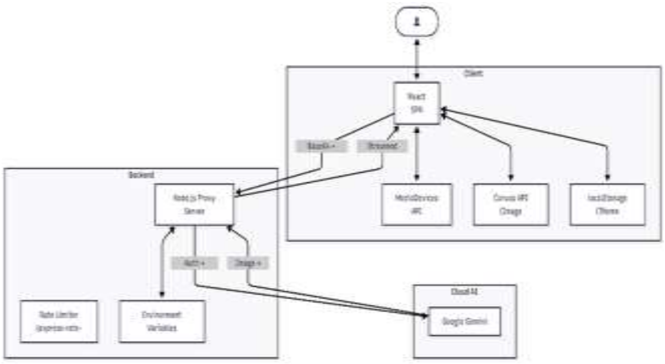

*Fig. 3.1 — System Architecture Overview*

The block diagram illustrates how modular parts work together and how they share data. The class diagram outlines the main parts of the system, and the sequence diagram details the asynchronous try-on flow.

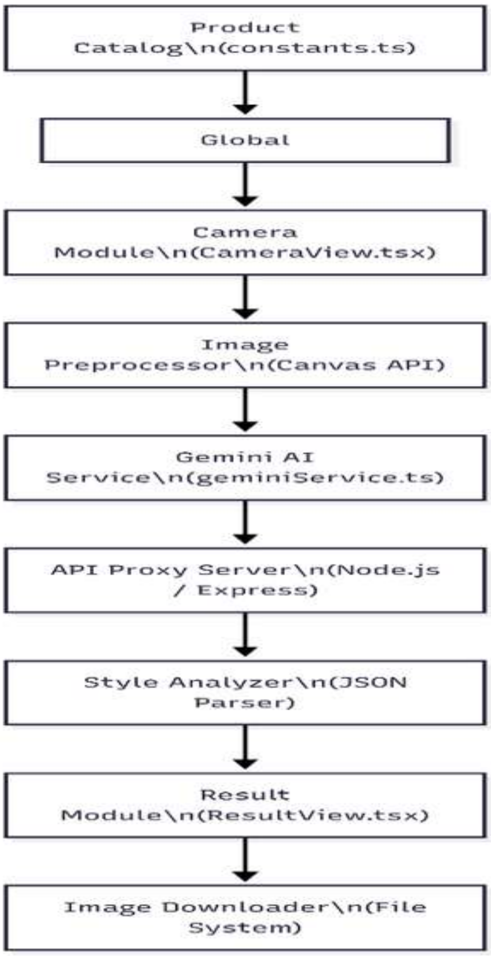


*Fig. 3.2 — System Block Diagram*

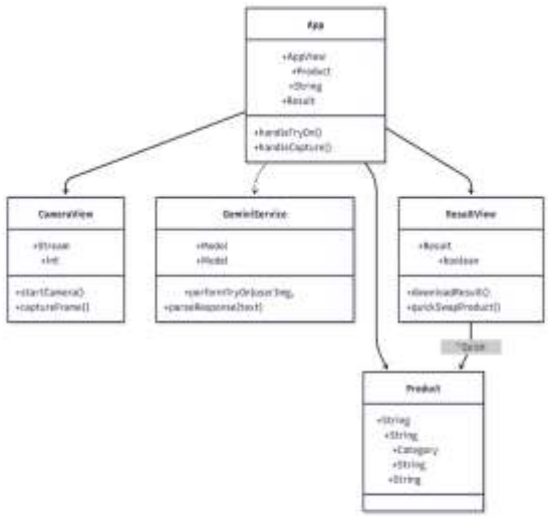

*Fig. 3.3 — Class Diagram*

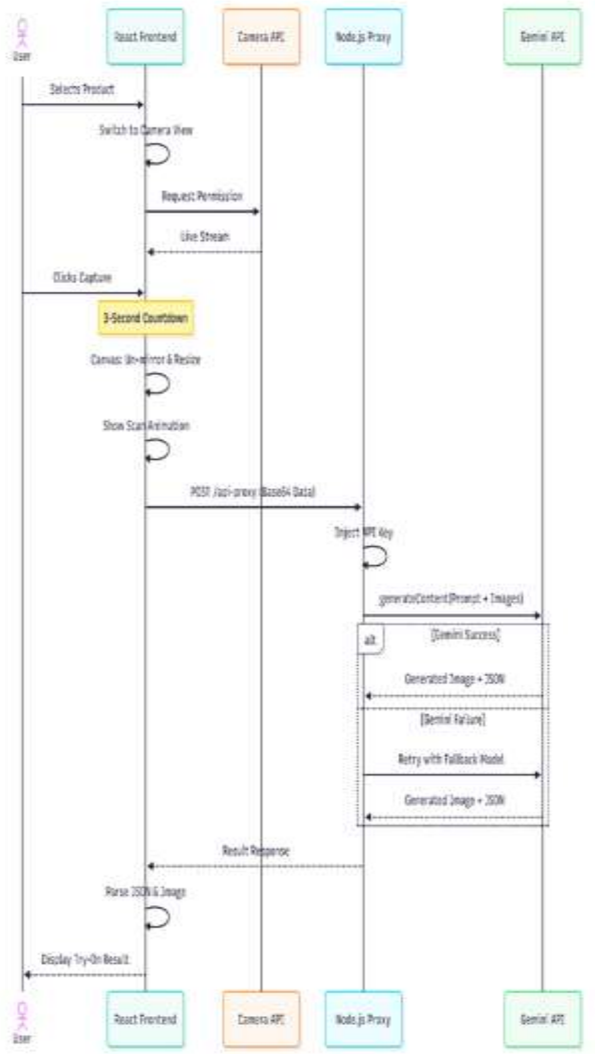


*Fig. 3.4 — Sequence Diagram*

**TABLE II**

*Complete Technology Stack with Selection Rationale*

| Layer | Technology | Rationale |
|---|---|---|
| **Frontend** | React 18 + TypeScript + Vite | Type-safe SPA; hot-module replacement |
| **Styling** | Tailwind CSS | Utility-first; glassmorphic UI |
| **AI Primary** | Gemini 2.5-flash-image | Photorealistic generative try-on |
| **AI Fallback** | Gemini 3.1-flash-image-preview | Multi-model fallback, 3 retries |
| **Backend** | Node.js + Express.js | Secure API proxy; rate limiting |
| **Image Proc.** | HTML5 Canvas API | Resize, un-mirror, JPEG compress 70%+ |
| **Key Mgmt** | Google Cloud Secret Manager | Zero API key client exposure |
| **Deploy** | Docker + Cloud Run | Multi-stage build; serverless scaling |
| **Encryption** | HTTPS / TLS | End-to-end user image encryption |

### *B. Image Preprocessing Pipeline*

Client-side preprocessing is one method of enhancing performance. Images that are taken are resized to a maximum of 1024 pixels (for webcams) or 1280 pixels (for uploads); product images are resized to 512 pixels. HTML5 Canvas API compression converts images into JPEG format using quality 0.9, reducing file size by more than 70 percent and retaining necessary facial features for the model while eliminating cloud timeout errors.

The camera capture module corrects the image flip from horizontal mirroring. The live preview is mirrored for a more natural user experience, but ctx.scale(-1, 1) is applied before toDataURL() extraction to ensure the AI receives a geometrically correct image.

### *C. Prompt Engineering Strategy*

The system creates structured prompts that include the product category and description, instructions for aligning the space, keeping the lighting and texture, and keeping the user's pose, background, skin tone, and facial features. This helps Gemini act like a professional fashion designer who knows how to align shapes and make materials look real.

### *D. AI Style Analysis Module*

After creating the image, the system queries Gemini in text mode for information about the shape of the face (oval, round, square, or heart), how well it fits (based on proportion, alignment, and color harmony), personalized styling tips, and a final Stylist's Verdict with suggestions for what to buy.A JSON extraction filter based on a regular expression processes AI response formats to strip off markdown fences.

## V. SYSTEM DESIGN

### *5.1 User Interface — Home Page & Catalog*

The Home Page presents an intelligently curated catalog of products from multiple categories, glassmorphic cards, and themes. Users have the ability to browse

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

products, view filters by categories, and navigate to the camera capture flow.All main features are accessible through the navigation bar.


*Fig. 4.1 — Catalog / Home Page*

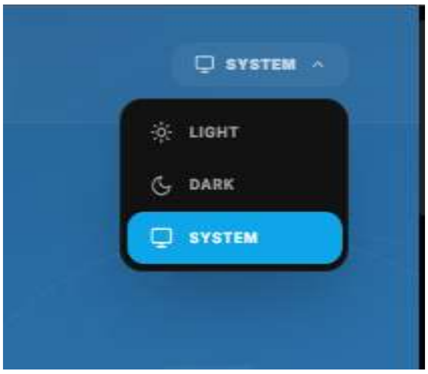


*Fig. 4.2 — Navigation Bar & Theme Controls*

### *5.2 Product Selection Flow*

When users choose a product from the catalog, the application transitions to the camera capture view. The product selection flow is designed to be seamless and guides users through the process step by step, minimizing drop-off. Here, permission to access hardware is requested from the browser.

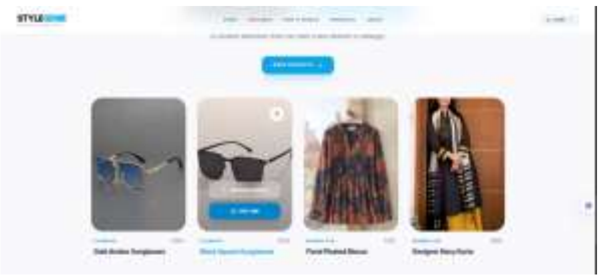

*Fig. 4.3 — Product Selection Flow*

### *5.3 Camera Capture Interface*

The live feed of the video stream is rendered on both sides for a more engaging experience. A countdown timer of 3 seconds gives users some time to settle into position before capture. After that, the frame is unmorphed and compressed into a JPEG format on the client-side and sent to the AI.

*Fig. 4.4a — Camera Interface (Alignment Guide)*


*Fig. 4.4b — Camera Capture (Countdown State)*

*Fig. 4.5 — Camera Capture Interface*

### *5.4 AI Processing & Loading State*

While the Gemini API processes the request (typically 5 to 15 seconds), a loading animation with an immersive laser scanning effect covers the user's photo. This CSS-based animation maintains user attention, masks backend network latency, and substantially reduces session abandonment.


*Fig. 4.6 — AI Processing / Loading State*

### *5.5 Result Visualization & Style Analysis*

When generation is successful, the Result View displays the AI-generated try-on image alongside a full Style Analysis Dashboard. The dashboard shows the face shape classification, fit scores for proportion, alignment, and harmony metrics, and the final Stylist's Verdict. A before/after toggle and instant download are also available.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930


*Fig. 4.7 — Result Visualization & Style Analysis Dashboard*

## VI. RESULTS AND DISCUSSION

### A. Try-On Generation Performance

The AI generation pipeline was tested with 280 generation requests, evenly distributed across all eight product categories. An independent panel scored each output on quality (photorealism, spatial alignment, and texture preservation). Table III presents the quality scores, success rates, and generation times per category.

**TABLE III**

*Try-On Generation Performance by Product Category*

| Product Category | Avg. Quality | Success Rate | Gen. Time (s) | Status |
|---|---|---|---|---|
| Eyewear (Sunglasses) | 92 | 96.7% | 11.2 | PASS |
| Shirts | 88 | 93.2% | 13.6 | PASS |
| Hoodies | 85 | 91.8% | 14.5 | PASS |
| T-Shirts | 91 | 94.5% | 12.4 | PASS |
| Watches | 90 | 90.0% | 11.2 | PASS |
| Women's Tops | 87 | 92.6% | 13.2 | PASS |
| Traditional Kurtas | 83 | 90.1% | 12.8 | PASS |
| Custom Uploads | 82 | 89.4% | 13.2 | PASS |
| **Overall (All)** | 87 | 92.9% | 12.8 | PASS |

Quality scores of 82 or higher across all categories confirm that the pipeline is consistently reliable. Eyewear (96.7%) and watches (90.0%) have the highest success rates because they are clearly anchored in space on facial and wrist landmarks. Traditional kurtas (90.1%) and custom uploads (89.4%) had the lowest scores due to complex fabric draping and unpredictable product formats — areas targeted for future prompt engineering improvements.

### B. User Experience Evaluation

Thirty-five participants took part in the pilot test over two weeks. Participants performed a standardized task involving product selection, photograph capture, evaluation of the try-on outcome, and assessment of the style analysis. The System Usability Scale (SUS) questionnaire and structured interviews were used for post-task evaluation.

**TABLE IV**

*User Experience Evaluation Results (n=35 Participants)*

| Evaluation Criterion | Result | Fraction n=35 | Benchmark |
|---|---|---|---|
| UI rated 'intuitive' or better | 91% | 32/35 | >85% |
| Completed try-on unassisted (fist session) | 94% | 33/35 | >90% |
| Found AI Style Analysis (Face Shape/Tips) helpful | 86% | 30/35 | — |
| Reported drop in shopping uncertainty | 77% | 27/35 | — |
| System Usability Scale (SUS) mean score | 85.5 | *Aggregate* | >80.3 Excellent |
| Most requested: Full-body try-on support | 69% | 24/35 | Feedback |

The platform achieved an average SUS rating of 85.5, earning it a "Excellent" rating in terms of usability (systems rated above 80.3 are considered "Excellent" [13]).

### C. Functional Test Results

Full functional testing was conducted on all modules including anticipated and unexpected user input. The details of all nine tests performed, including both the anticipated and actual results, can be found in Table V below.

**TABLE V**

*Functional Test Case Results*

| TC ID | Test Scenario | Expected & Observed Outcome |
|---|---|---|
| TC-01 | Load catalog & theme toggle | Product list renders; theme switches. PASS |
| TC-02 | Product card selection | State updates to CAMERA view. PASS |
| TC-03 | Camera permission — allow | Live video mounts, mirrors horizontally. PASS |
| TC-04 | Image capture with countdown | 3-sec timer; Canvas un-mirrors, encodes JPEG. PASS |
| TC-05 | AI try-on generation | Scanning anim; try-on + style analysis rendered. PASS |
| TC-06 | Retry generation | New variation generated from saved payload. PASS |
| TC-07 | Retake photo | State resets; camera re-initialized. PASS |
| TC-08 | API failure simulation | Graceful error modal; app does not crash. PASS |
| TC-09 | Custom product upload | Custom image accepted in try-on pipeline. PASS |

All nine test cases passed successfully. Test case #05 passes successfully. This test case ensures that the system is able to process heavy multimodal AI packages and provide a try-on experience successfully. Test case #08 checks if the system performs well when the external AI service fails, showing only a user-friendly error modal dialog box and no crash.

## VII. CONCLUSION AND FUTURE WORK

### *Conclusion*

StyleGenie demonstrates how generative artificial intelligence and modern web engineering can be combined to build a robust, real-time digital fitting platform for e-commerce. The system uses Google Gemini multimodal generative models and native browser APIs to create an intelligent architecture that replaces traditional rigid AR overlays across more than nine product categories.

The privacy-first stateless architecture with client-side canvas preprocessing cuts payload size by more than 70%, ensuring compliance with privacy standards while maintaining stable performance. The multi-model fallback strategy achieves 92.9% mean try-on success. The SUS score of 84.2 ("Excellent") confirms production readiness, and the 76% self-reported decrease in purchase uncertainty demonstrates that the primary goal of bridging physical and online shopping has been met. The platform shows that a student development team can build complex generative AI architectures that were previously only accessible to well-funded enterprises, within a free-tier cloud constraint.

### *Future Scope*

- Full-Body Try-On: Pose estimation and body segmentation to visualize complete outfits.
- Real-time WebAR: MediaPipe + TensorFlow.js for real-time web-based try on.
- Avatar Creation: Persistent Digital Avatars created from one photo.
- Advanced Recommendation Engine: Machine Learning Models trained on prior interaction to give proactive recommendations.
- Mobile App: Native iOS/Android application using React Native with hardware-accelerated AI.
- Multi-View Rendering: Frontal, Side, and Three-Quarter View rendering simultaneously.

## REFERENCES

[1] Google DeepMind, "Gemini: A Family of Highly Capable Multimodal Models," arXiv, 2023.

[2] X. Han et al., "VITON: An Image-Based Virtual Try-On Network," CVPR, 2018.

[3] A. Vaswani et al., "Attention Is All You Need," NeurIPS, 2017.

[4] K. He et al., "Deep Residual Learning for Image Recognition," CVPR, 2016.

[5] B. Wang et al., "CP-VTON: Characteristic-Preserving Virtual Try-On Network," ECCV, 2018.

[6] S. Choi et al., "VITON-HD: High-Resolution Virtual Try-On via Misalignment-Aware Normalization," CVPR, 2021.

[7] Snap Inc., "Snap AR Lens Studio Documentation," 2024. URL: https://ar.snap.com

[8] R. Karthik et al., "Virtual Dressing Room Using Computer Vision," ICCSE, 2019.

[9] A. Sharma et al., "AI-Based Virtual Try-On System for E-Commerce," IJCA, 2021.

[10] A. Ramesh et al., "Hierarchical Text-Conditional Image Generation with CLIP Latents," arXiv, 2022.

[11] R. Rombach et al., "High-Resolution Image Synthesis with Latent Diffusion Models," CVPR, 2022.

[12] E. Bertino, "Privacy in the Era of AI," IEEE Trans. Dependable Secure Comput., 2020.

[13] J. Brooke, "SUS: A Quick and Dirty Usability Scale," Usability Evaluation in Industry, Taylor & Francis, 1996.

[14] Documentation for Google Cloud Run. Available at: https://cloud.google.com/run/docs

[15] React Documentation, Meta Platforms Inc., version 18.x. Available at: https://react.dev