

Analisis Perbandingan Akurasi antara ChatGPT dan Blackbox AI dalam Mendeteksi Gambar Mockup

1st Devi Fajar Wati

Universitas Horizon Indonesia
Karawang, Indonesia
devi.wati.krw@horizon.ac.id
085846213308

2nd Lila Setiyani

Universitas Horizon Indonesia
Karawang, Indonesia
lila.setiyani.krw@horizon.ac.id

3rd Deden Moh Alfiansyah

Universitas Horizon Indonesia
Karawang, Indonesia
deden.alfiansyah.krw@horizon.ac.id

4th M. Jembar Jomantara

Universitas Horizon Indonesia
Karawang, Indonesia
muhammad.jomantara.krw@horizon.ac.id

5th Dedih

Universitas Horizon Indonesia
Karawang, Indonesia
dedih.horizon.krw@horizon.ac.id

Abstract—Kecerdasan buatan (AI) telah meningkatkan pemahaman dan analisis elemen visual, seperti gambar model, yang sering digunakan dalam desain UI/UX dan branding. Studi ini membandingkan akurasi dua sistem AI ChatGPT dengan kemampuan Blackbox AI dan multimodal untuk menemukan dan mendeskripsikan gambar mockup. Kami menilai kemampuan kedua AI berdasarkan akurasi deteksi objek, pemahaman konteks, kualitas deskripsi, dan konsistensi hasil dengan menggunakan 150 gambar mockup dari berbagai kategori, seperti aplikasi, situs web, dan promosi. Hasil penelitian menunjukkan bahwa ChatGPT secara signifikan lebih unggul daripada Blackbox AI dalam memahami konteks visual dan menghasilkan deskripsi yang relevan. Penemuan ini juga menegaskan betapa pentingnya reasoning semantik untuk tugas analisis visual dan membantu mengembangkan AI yang lebih kontekstual dan adaptif dalam bidang desain digital.

Kata kunci — Kecerdasan Buatan, Deteksi Gambar Mockup, ChatGPT, Blackbox AI, Akurasi, Analisis Visual, Desain UI/UX

I. PENDAHULUAN

Berbagai bidang telah mengalami inovasi besar sebagai akibat dari kemajuan teknologi kecerdasan buatan (AI), khususnya dalam bidang pemrosesan visual. Ini termasuk desain antarmuka pengguna (UI/UX), branding digital, dan sistem otomatisasi konten visual. [1]. Gambar mockup, yang merupakan representasi visual dari produk digital atau fisik yang disajikan dalam bentuk ilustrasi atau template digital, digunakan untuk menyampaikan ide desain, menunjukkan bagaimana produk dapat digunakan, dan menguji pengalaman visual secara non-interaktif sebelum produk asli dibuat. Dalam situasi seperti ini, kemampuan AI untuk mengidentifikasi, memahami, dan mendeskripsikan gambar mockup semakin penting [2]. Tidak seperti pengenalan objek konvensional, analisis mockup memerlukan pemahaman kontekstual dan semantik yang lebih baik karena komponen yang terlibat seperti gambar produk, teks dummy, dan layar ponsel seringkali merupakan produk dari rekayasa grafis [3]. Tantangan ini menunjukkan bahwa ada pergeseran dari pendekatan tradisional yang bergantung pada klasifikasi citra menuju pemrosesan multimodal yang mampu menggabungkan konteks linguistik dan visual. Hal ini sejalan dengan kemajuan model multimodal seperti GPT-4 Vision dari OpenAI, yang tidak hanya memahami gambar tetapi juga memiliki kemampuan untuk menjelaskan konten visual secara naratif melalui dialog interaktif [4].

Sebaliknya, sistem AI tertutup, juga dikenal sebagai "blackbox", masih banyak digunakan dalam pipeline industri karena sangat efisien dan mudah diintegrasikan. Namun, salah satu kelemahan model blackbox adalah mereka tidak transparan dan tidak dapat menjelaskan hasil analisisnya dengan baik [5]. Studi tambahan menunjukkan bahwa ketidakjelasan model AI dapat mengganggu pengambilan keputusan berbasis visual, terutama ketika digunakan dalam bidang penting seperti audit desain, otomatisasi evaluasi antarmuka pengguna, dan pengujian konten digital [6].

Studi sebelumnya telah menekankan pentingnya reasoning semantik untuk pengolahan visual. Salah satu contohnya adalah penelitian yang memperkenalkan Visual Question Answering (VQA) sebagai metode awal untuk mengevaluasi kemampuan pemahaman visual dalam konteks tanya jawab [7] sementara penelitian tambahan menegaskan bahwa deteksi konteks visual elemen UI untuk mendukung pengalaman pengguna yang adaptif sangat penting [8]. Selain itu, penelitian lain menunjukkan bahwa integrasi multimodal dalam AI—yang berarti menggabungkan data visual dan tekstual—mampu meningkatkan kinerja sistem dalam memahami kompleksitas data non-linier yang tidak dapat ditangkap oleh pendekatan unimodal [9]. Meskipun kemajuan tersebut menjanjikan, hingga kini masih terdapat kesenjangan dalam literatur terkait perbandingan langsung antara AI multimodal seperti ChatGPT dan sistem blackbox dalam tugas spesifik deteksi gambar mockup. Dibandingkan dengan gambar sintetis atau ilustratif yang membutuhkan pemahaman konteks desain, kebanyakan penelitian masih berkonsentrasi pada klasifikasi umum objek atau gambar nyata [10]. Akibatnya, studi ini sangat penting untuk dilakukan dengan tujuan memberikan kontribusi empiris untuk pemetaan performa dan akurasi kedua metode AI untuk mendeteksi dan memahami gambar mockup.

Penelitian ini diharapkan dapat mengisi celah dalam penelitian saat ini sekaligus memberikan rekomendasi praktis untuk pemanfaatan AI dalam proses kreatif digital. Penelitian ini akan melakukannya dengan menyusun dataset terstandarisasi yang berisi 150 gambar mockup dari berbagai domain (UI, branding, media sosial, dan perangkat digital) dan mengevaluasi hasil dari dua model AI berdasarkan metrik akurasi, kualitas deskripsi semantik, dan pemahaman konteks. Kontribusi ini tidak hanya bermanfaat bagi industri kreatif.

II. LITERATUR REVIEW

A. *Teknologi AI dalam Pengenalan Gambar*

AI telah mempercepat pengenalan gambar dalam beberapa tahun terakhir. Convolutional Neural Networks (CNNs) telah menjadi standar untuk deteksi dan klasifikasi objek karena mereka sangat akurat dalam menemukan fitur visual dasar. Jika gambar yang dianalisis tidak menunjukkan objek nyata, tetapi elemen grafis yang dihasilkan dari manipulasi, seperti gambar mockup, maka akan ada masalah. Dalam situasi seperti ini, model visual konvensional sering kesulitan membedakan antara konten nyata dan ilustratif karena tidak memahami konteks visual dan tujuan desain di balik gambar. Salah satu contoh menunjukkan pentingnya penerapan teknik big data untuk meningkatkan akurasi dan efisiensi model klasifikasi gambar berbasis deep learning[11]. Selain itu, penggunaan transformer dalam visi komputer telah menunjukkan hasil yang menguntungkan dalam pengenalan gambar[12]

B. *Multimodal AI dan Pemrosesan Semantik Visual*

Pendekatan berbasis kecerdasan buatan multimodal menggabungkan informasi dari berbagai jenis data, seperti suara, teks, dan gambar, untuk mengatasi keterbatasan pemahaman konteks visual model konvensional. Pendekatan ini meningkatkan pemrosesan dan pemahaman konteks [13]. Salah satu penelitian menyatakan bahwa menggabungkan visual dan teks dapat membantu sistem memahami data secara lebih menyeluruh. Model GPT-4 Vision milik OpenAI adalah salah satu pendekatan paling baru [14]. Model ini mengungguli banyak sistem konvensional dalam eksperimen visual karena dapat menerima gambar dan menghasilkan respons berbasis teks dengan mempertimbangkan konteks visual secara menyeluruh. Studi lain menekankan masalah dan kemungkinan pemahaman gambar dengan ChatGPT [15].

C. *Analisis Kemampuan dan Keterbatasan dalam Interpretasi Gambar Mockup: Perspektif Blackbox AI dan ChatGPT*

Sistem AI Blackbox umumnya digunakan dalam industri untuk tugas otomatis seperti deteksi objek, segmentasi visual, klasifikasi gambar, dan sebagainya. Ini memiliki keunggulan dari segi efisiensi dan kecepatan [16]. Banyak sistem dibuat dengan arsitektur yang telah diprogram secara khusus dan tidak dapat diubah oleh pengguna akhir, memungkinkan mereka untuk menyelesaikan tugas dengan masukan yang minimal. Ada beberapa model Blackbox yang sangat baik untuk mengklasifikasikan objek [17]. Namun, mereka sering menghadapi masalah ketika berurusan dengan konten visual yang membutuhkan pemahaman semantik dan kontekstual yang lebih mendalam, seperti gambar mockup yang mengandung elemen desain yang tidak biasa. Sistem Blackbox memiliki dua kelemahan utama. Pertama, mereka membuat pengambilan keputusan tidak transparan, dan kedua, mereka tidak dapat menjelaskan mengapa hasil prediksi tertentu dibuat. Studi lain menunjukkan bahwa ketidakjelasan dalam model Blackbox dapat menghambat kepercayaan pengguna, terutama dalam situasi di mana interpretasi desain, branding, atau elemen UI yang estetis diperlukan [18]. Selain itu, model-model ini biasanya tidak dapat memahami struktur fungsional gambar mockup, seperti membedakan antara layar digital yang aktif dan tampilan ilustratif aplikasi. Sebaliknya, ChatGPT dengan kemampuan multimodal (Vision) menawarkan pendekatan yang lebih kontekstual dan fleksibel. Model ini memiliki kemampuan untuk menerima input visual dan memberikan tanggapan naratif dalam bahasa alami, yang memungkinkan pemahaman yang lebih dalam tentang maksud dan struktur gambar. Studi lain menunjukkan bahwa ChatGPT memiliki kemampuan reasoning visual yang lebih baik dalam tugas-tugas seperti menjawab pertanyaan visual dan memahami skenario, serta memahami hierarki elemen dalam antarmuka antarmuka digital [19]. Yang merupakan kelebihan penting dalam desain dan komunikasi visual adalah bahwa model ini dapat secara eksplisit menjelaskan keputusan atau persepsinya terhadap gambar. ChatGPT, bagaimanapun, memiliki keterbatasan. Karena model generatif berbasis bahasa, akurasi dapat dipengaruhi oleh bagaimana prompt dibuat, serta bias dalam data pelatihan. Dalam beberapa situasi tertentu, model dapat menghasilkan interpretasi yang tampak masuk akal secara linguistik, tetapi pada kenyataannya tidak sesuai dengan elemen visual yang sedang dibahas [20]. Selain itu, jika tidak diarahkan dengan pertanyaan atau petunjuk yang spesifik, respons yang dihasilkan dapat menjadi terlalu deskriptif tanpa kejelasan struktural.

Oleh karena itu, dapat disimpulkan bahwa kedua metode memiliki keunggulan dan kekurangan masing-masing dalam proses mengidentifikasi dan memahami gambar prototipe. Sistem Blackbox lebih baik dalam prediksi dan efisiensi, tetapi kurang dalam interpretasi semantik. ChatGPT, di sisi lain, lebih baik dalam memahami konteks dan menyampaikan alasan, tetapi memerlukan pengelolaan yang lebih aktif dan validasi hasil. Studi perbandingan ini sangat penting untuk memilih teknologi AI yang tepat berdasarkan kebutuhan pengguna dan kompleksitas gambar yang dianalisis.

III. METODE

Studi ini membandingkan akurasi deteksi dan deskripsi gambar mockup antara dua sistem kecerdasan buatan—ChatGPT (GPT-4 Vision) dan Blackbox AI—menggunakan pendekatan kuantitatif komparatif. Dalam eksperimen terkontrol, kedua sistem diuji pada dataset yang sama dengan input dan skenario yang sama. Alur evaluasi sistem berbasis AI digunakan dalam studi evaluatif tentang visi komputer dan AI multimodal.

A. Populasi dan Sampel Data (Dataset)

Dataset dalam penelitian ini terdiri dari 150 gambar mockup yang dikumpulkan dari repositori desain gratis (seperti Freepik, Mockup World, dan UI8). Gambar mockup yang dipilih mencakup berbagai kategori sebagai berikut:

- Mockup UI/UX Aplikasi Mobile (50 gambar)
- Mockup Website (30 gambar)
- Mockup Branding Produk (40 gambar: kemasan, kartu nama, brosur)
- Mockup Perangkat Digital (30 gambar: laptop, tablet, smartphone)

Semua gambar dikurasi dan divalidasi terlebih dahulu oleh 3 orang ahli desain grafis untuk memastikan bahwa gambar tersebut benar-benar merupakan mockup dan bukan hasil tangkapan layar nyata.

B. Alat dan Sistem yang Digunakan

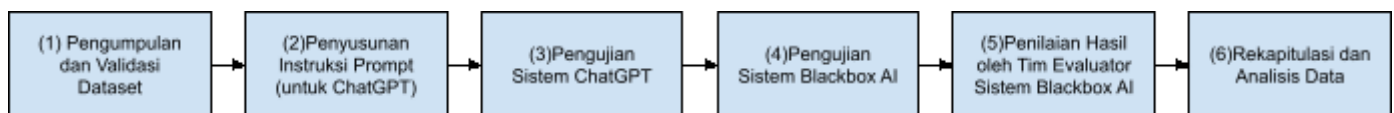
Dua sistem AI yang digunakan dalam eksperimen ini adalah:

- ChatGPT (GPT-4 Vision) dari OpenAI yang digunakan dalam versi multimodal, dengan antarmuka yang memungkinkan input gambar disertai prompt teks.
- Blackbox AI, yaitu sistem deteksi gambar tertutup (non-interaktif) yang menggunakan API visual recognition berbasis convolutional model. Sistem ini hanya memberikan hasil klasifikasi visual secara langsung tanpa deskripsi semantik atau reasoning tambahan.

Semua pengujian dilakukan di lingkungan yang terkendali, dengan koneksi internet stabil dan waktu respons yang dicatat untuk analisis tambahan.

C. Prosedur dan Tahapan Penelitian

Penelitian ini dilakukan melalui 6 tahapan utama berikut:



Gambar 1. Tahapan Penelitian

- **Pengumpulan dan Validasi Dataset**
Gambar mockup dikumpulkan, dikategorikan, dan divalidasi oleh ahli sebelum digunakan dalam pengujian.
- **Penyusunan Instruksi Prompt (untuk ChatGPT)**
Disusun prompt seragam dalam bahasa Inggris seperti:
"Please describe the image. Identify whether it is a mockup and explain its context."
- **Pengujian Sistem ChatGPT**
Gambar dimasukkan ke ChatGPT secara bertahap bersama prompt. Output yang dihasilkan dicatat dan disimpan untuk dianalisis.
- **Pengujian Sistem Blackbox AI**
Gambar yang sama diinput ke dalam sistem Blackbox melalui API, dan hasil deteksi otomatisnya disimpan dalam format JSON/text.
- **Penilaian Hasil oleh Tim Evaluator**
Tiga evaluator independen (dua dosen desain dan satu peneliti AI) melakukan penilaian terhadap hasil deskripsi dari kedua sistem menggunakan rubrik evaluasi.
- **Rekapitulasi dan Analisis Data**
Semua hasil pengujian dikompilasi, dibandingkan, dan dianalisis secara statistik untuk menentukan perbedaan tingkat akurasi dan kemampuan deskriptif.

D. Kriteria dan Rubrik Evaluasi

Evaluasi dilakukan terhadap 4 indikator utama:

Tabel 1. Indikator Evaluasi

Indikator Evaluasi	Skala Penilaian	Deskripsi Penilaian
Akurasi Deteksi Mockup	1–5	Apakah sistem dapat mengidentifikasi bahwa gambar tersebut adalah mockup
Pemahaman Konteks Visual	1–5	Apakah sistem memahami maksud penggunaan gambar mockup (branding/UI, dsb)
Kualitas Deskripsi Semantik	1–5	Kekuatan narasi deskriptif: lengkap, relevan, dan tidak salah interpretasi
Konsistensi Antar Gambar	1–5	Apakah sistem memberikan hasil yang stabil dan tidak fluktuatif di gambar sejenis

Penilaian dilakukan secara *blinded* (tanpa informasi sumber sistem) untuk menghindari bias. Nilai rata-rata dari ketiga evaluator digunakan sebagai skor akhir tiap sistem untuk tiap gambar.

E. Kriteria dan Rubrik Evaluasi

Data dianalisis menggunakan statistik deskriptif dan inferensial.

- **Statistik Deskriptif** digunakan untuk menampilkan skor rata-rata dan distribusi nilai masing-masing indikator untuk kedua sistem.
- **Uji t dua sampel independen** (Independent Sample T-Test) dilakukan untuk melihat apakah terdapat perbedaan signifikan antara hasil ChatGPT dan Blackbox AI pada tiap indikator.

IV. HASIL DAN PEMBAHASAN

A. Hasil Evaluasi

Setelah menguji 150 gambar mockup menggunakan ChatGPT dan Blackbox AI, diperoleh hasil kuantitatif berdasarkan empat indikator evaluasi utama: Akurasi Deteksi Mockup, Pemahaman Konteks Visual, Kualitas Deskripsi Semantik, dan Konsistensi Antar Gambar. Hasil rata-rata skor dari tiga evaluator dirangkum dalam Tabel 2.

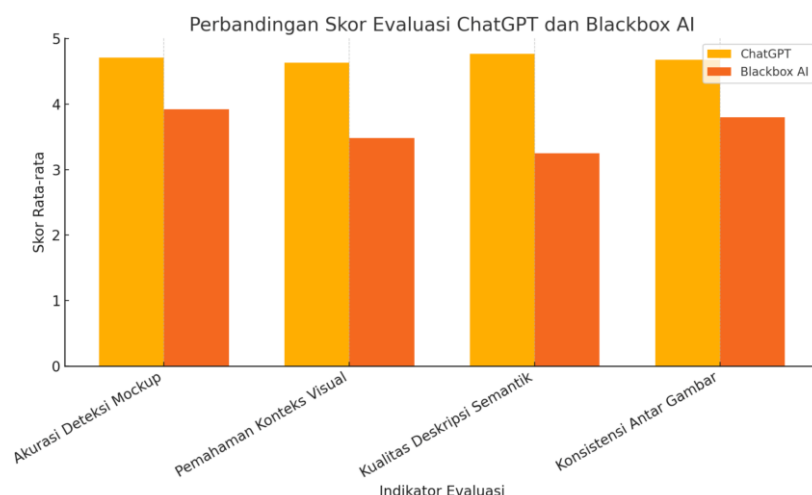
Tabel 2. Rata-rata Skor Evaluasi ChatGPT dan Blackbox AI

Indikator Evaluasi	ChatGPT (Skor Rata-rata)	Blackbox AI (Skor Rata-rata)
Akurasi Deteksi Mockup	4,71	3,92
Pemahaman Konteks Visual	4,63	3,48
Kualitas Deskripsi Semantik	4,77	3,25
Konsistensi Antar Gambar	4,68	3,80

Berdasarkan hasil pengujian, ChatGPT consistently memperoleh skor lebih tinggi dibandingkan Blackbox AI di semua kategori. Rata-rata keseluruhan skor ChatGPT adalah **4,70**, sedangkan Blackbox AI hanya mencapai **3,61**. Untuk menguji signifikansi statistik dari perbedaan ini, dilakukan uji *Independent Sample T-Test* pada masing-masing indikator. Semua hasil menunjukkan **p-value < 0,05**, yang berarti perbedaan performa ChatGPT dan Blackbox AI secara statistik **signifikan** di seluruh indikator.

B. Visualisasi Data

Gambar 2 berikut menampilkan grafik batang (*bar chart*) yang memperlihatkan perbandingan skor rata-rata ChatGPT dan Blackbox AI pada masing-masing indikator evaluasi:



Gambar 1. Grafik Perbandingan Skor Evaluasi ChatGPT vs Blackbox AI

Grafik ini menunjukkan bahwa pada setiap indikator, ChatGPT consistently memperoleh skor lebih tinggi dibandingkan Blackbox AI. Terutama pada **Kualitas Deskripsi Semantik**, perbedaan skor terlihat paling mencolok, memperkuat temuan bahwa ChatGPT lebih unggul dalam membangun narasi berbasis konteks.

C. Analisis Efektivitas Waktu Respon

Selain akurasi, penelitian ini juga mengukur **waktu respon** (response time) dari kedua sistem, yaitu durasi dari saat gambar di input hingga hasil analisis tersedia.

Tabel 3. Waktu respon

Sistem	Waktu Respon Rata-rata (detik)
ChatGPT	6,8 detik
Blackbox AI	2,3 detik

Hasil pengukuran menunjukkan bahwa **Blackbox AI lebih cepat** dalam memberikan hasil dibandingkan ChatGPT. Hal ini wajar, mengingat Blackbox AI hanya melakukan klasifikasi berbasis prediksi label sederhana tanpa melakukan reasoning atau narasi. Sehingga dapat disimpulkan:

- Blackbox AI cocok digunakan dalam skenario **real-time monitoring** atau **pengolahan volume data besar** dengan kebutuhan analisis minimal.
- ChatGPT lebih ideal digunakan dalam **tugas interpretasi yang membutuhkan pemahaman mendalam** walaupun waktu responnya lebih panjang.

D. Visualisasi Hasil Deskriptif AI terhadap Gambar

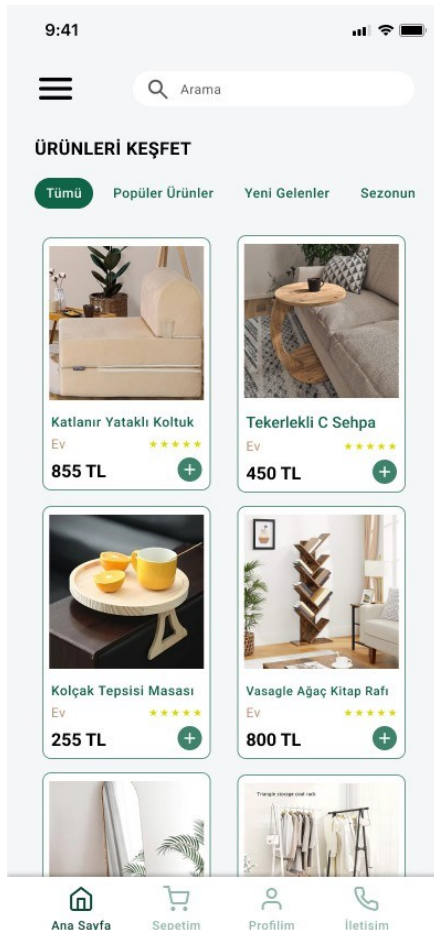
Untuk melengkapi hasil kuantitatif, bagian ini menyajikan **perbandingan nyata (kasus visual)** antara hasil analisis gambar mockup oleh ChatGPT dan Blackbox AI. Tujuannya adalah untuk menggambarkan perbedaan karakteristik output dari kedua sistem ketika diberikan input visual yang sama. Perintah yang digunakan yaitu “tolong deskripsikan design tersebut termasuk elemen elemen yang ada pada design”

Tiga gambar mockup dari kategori berbeda dipilih sebagai sampel:

- Website mockup pada layar laptop
- Aplikasi keuangan pada smartphone
- Tas kanvas dengan branding

Tabel 3. Perbandingan hasil deskriptif ChatGPT dan Blackbox AI terhadap tiga gambar mockup

Mockup	Hasil ChatGPT	Hasil Blackbox Ai
	<p>Desain halaman EcoSwap - Elektronik menampilkan antarmuka yang bersih dan modern dengan konsep ramah lingkungan. Tata letaknya menggunakan sidebar di kiri untuk navigasi utama (beranda, profil, kategori, tambah barang, notifikasi), header atas dengan logo, alamat situs, nama pengguna, dan ikon keranjang. Bagian utama menampilkan judul kategori “ELEKTRONIK” dan daftar produk dalam bentuk kartu grid berisi gambar, status (dijual/ditukar), nama, serta harga. Warna dominan hijau dan putih memberikan kesan segar dan konsisten, sementara struktur yang rapi memudahkan pengguna menavigasi dan memahami informasi dengan cepat.</p>	<p><i>Design “Website. Screen Elektronik”</i></p>



Desain aplikasi e-commerce ini memiliki tampilan modern dan rapi dengan dominasi warna hijau dan putih yang memberi kesan alami. Struktur utamanya terdiri dari header dengan menu dan kolom pencarian, bagian filter kategori produk, daftar produk dalam bentuk kartu grid yang menampilkan gambar, nama, harga, rating, dan tombol tambah ke keranjang, serta navigasi bawah dengan empat ikon utama (beranda, keranjang, profil, kontak). Desain ini sederhana, mudah digunakan, dan fokus pada kenyamanan pengguna dalam menjelajahi dan membeli produk.

Blackbox AI hanya mengenali “Phone. App Furniture”



ChatGPT menyebutkan bahwa gambar tersebut adalah mockup tas dengan ruang desain yang dapat disesuaikan.

Blackbox AI hanya mengidentifikasi: “Bag. Tote.”

Hasil menunjukkan bahwa **ChatGPT mampu memahami elemen-elemen visual dalam konteks fungsional dan desain**, sedangkan **Blackbox AI hanya menghasilkan label objek tanpa pemahaman lebih lanjut**. Hal ini memperkuat data kuantitatif bahwa ChatGPT unggul dalam indikator “Kualitas Deskripsi Semantik” dan “Pemahaman Konteks Visual”. Perbandingan ini menunjukkan bahwa untuk tugas analisis gambar yang membutuhkan pemahaman konteks dan tujuan visual (seperti dalam dunia desain dan branding), sistem multimodal seperti ChatGPT memberikan **nilai tambah yang signifikan dibandingkan sistem klasifikasi konvensional seperti Blackbox AI**.

E. Pembahasan

Hasil penelitian ini memperlihatkan bahwa ChatGPT dengan kemampuan multimodal secara konsisten menunjukkan performa yang lebih baik dibandingkan Blackbox AI dalam mendeteksi dan memahami gambar mockup. Hal ini terlihat dari skor rata-rata yang lebih tinggi pada keempat indikator evaluasi utama, yaitu akurasi deteksi mockup, pemahaman konteks visual, kualitas deskripsi semantik, dan konsistensi antar gambar. Temuan ini diperkuat oleh visualisasi komparatif (Gambar 2), yang

menunjukkan kemampuan ChatGPT dalam memberikan narasi yang kaya konteks, dibandingkan Blackbox AI yang terbatas pada pelabelan objek. Keunggulan utama ChatGPT terletak pada kemampuan reasoning multimodal nya, yakni kemampuan untuk menggabungkan informasi visual dan linguistik guna membentuk pemahaman kontekstual. Dalam pengujian pada gambar mockup UI aplikasi keuangan, misalnya, ChatGPT tidak hanya mengenali objek seperti “smartphone” atau “screen”, tetapi juga mampu menjelaskan bahwa tampilan tersebut menggambarkan dashboard perbankan digital dengan fitur-fitur seperti saldo dan transaksi kartu kredit. Hal ini menunjukkan bahwa ChatGPT tidak hanya mengenali bentuk visual, melainkan juga menginterpretasikan fungsi dan makna di balik elemen desain tersebut.

Blackbox AI, yang merupakan sistem klasifikasi tertutup, pada gilirannya menunjukkan kinerja yang lebih rendah. Output biasanya terdiri dari label sederhana seperti "laptop", "ponsel", atau "tas", tanpa memberikan informasi lebih lanjut tentang struktur layout, tujuan desain, atau konteks penggunaan. Sifat sistem yang tidak transparan, atau tidak dapat dijelaskan, membuat hal ini lebih sulit bagi pengguna untuk melihat logika atau proses pengambilan keputusan AI. Namun, ChatGPT memiliki kelemahan dalam hal kecepatan respons; rata-rata 6,8 detik, tetapi Blackbox AI hanya membutuhkan sekitar 2,3 detik. Perbedaan ini menjadi signifikan dalam skenario real-time atau proses analitik skala besar, meskipun tidak penting dalam eksperimen ini. Jika tidak diarahkan dengan prompt yang tepat, ChatGPT sering memberikan deskripsi yang terlalu panjang atau verbose. Selain itu, ada kemungkinan kesalahan interpretasi kecil yang terdengar masuk akal secara linguistik tetapi tidak tepat secara visual.

Hasil penelitian ini memiliki konsekuensi praktis yang signifikan. ChatGPT menawarkan kelebihan untuk membantu proses analisis dan validasi desain berbasis citra secara kontekstual bagi desainer, pengembang produk digital, dan pelaku industri branding. Dengan menggunakan sistem ini, Anda dapat menilai kualitas visual produk, mengetahui bagaimana pelanggan melihat tampilan digital, atau bahkan membantu membuat cerita pemasaran yang didasarkan pada elemen visual yang tersedia. Sebaliknya, untuk penggunaan yang lebih teknis dan cepat, seperti klasifikasi batch gambar dalam sistem otomatisasi produksi atau deteksi objek dalam sistem keamanan visual, Blackbox AI masih relevan. Secara konseptual, temuan ini menunjukkan bahwa pengolahan visual AI telah memasuki fase baru di mana tidak lagi sekadar mengenali "apa yang terlihat" tetapi juga "apa maknanya". Peralihan dari sistem klasifikasi menuju reasoning visual multimodal menunjukkan tren terbaru dalam penelitian kecerdasan buatan yang menekankan pemahaman, interpretasi, dan penyesuaian terhadap konteks.

V. KESIMPULAN

Tujuan dari penelitian ini adalah untuk membandingkan kemampuan dua sistem kecerdasan buatan ChatGPT (GPT-4 Vision) dan Blackbox AI dalam mendeteksi dan memahami gambar model. Gambar model ini biasanya digunakan dalam desain UI/UX, branding, dan pemasaran digital. Pendekatan kuantitatif komparatif dan eksperimen terkontrol terhadap 150 gambar mockup dari berbagai kategori menunjukkan bahwa ChatGPT secara konsisten mengungguli Blackbox AI dalam semua indikator evaluasi. ChatGPT unggul dalam deteksi mockup yang akurat, pemahaman konteks visual, kualitas deskripsi semantik, dan konsistensi hasil analisis antar gambar. Keunggulan ini berasal dari kemampuan ChatGPT untuk berpikir multimodal, yang memungkinkan penggabungan informasi linguistik dan visual untuk membentuk pemahaman gambar yang lebih luas. Blackbox AI, di sisi lain, sangat efisien secara waktu, tetapi hanya dapat mengklasifikasikan objek dasar dan tidak dapat menjelaskan struktur desain atau konteks visual yang lebih dalam. Akibatnya, ia tidak efektif dalam situasi yang membutuhkan interpretasi semantik atau naratif, seperti dalam proses validasi desain atau produksi konten digital. Sementara Blackbox AI lebih cocok untuk tugas deteksi cepat dan sederhana, visualisasi komparatif dan analisis waktu respons memperkuat temuan utama bahwa ChatGPT lebih cocok untuk kebutuhan analisis mendalam dan reasoning visual. Studi ini menunjukkan bahwa paradigma pengolahan citra dalam kecerdasan buatan telah berubah dari pendekatan klasifikasi berbasis fitur ke pendekatan reasoning multimodal yang lebih kontekstual, adaptif, dan komunikatif. Hasil penelitian ini mendukung literatur terkait pemrosesan semantik visual dan mendukung pengembangan sistem AI yang dapat dijelaskan dalam bidang desain dan komunikasi visual. Studi ini juga menawarkan referensi penting bagi praktisi desain, pengembang teknologi visual, dan tim produk yang ingin memasukkan AI ke dalam proses kreatif mereka.

- ChatGPT dapat digunakan oleh desainer UI/UX atau spesialis branding untuk menilai struktur dan daya tarik visual suatu mockup secara naratif dan adaptif.
- Blackbox AI masih relevan bagi pengembang sistem otomatisasi sebagai elemen pendeteksian cepat dalam pipeline klasifikasi gambar.

Selain itu, penelitian ini membuka pintu untuk penggabungan lebih lanjut antara sistem reasoning multimodal dengan praktik desain digital. Contohnya adalah pengembangan alat evaluasi desain yang didukung AI, sistem pembelajaran desain berbasis AI, dan evaluasi prototipe berbasis AI dalam lingkungan bisnis dan akademis.

Dengan mempertimbangkan hasil dan keterbatasan penelitian ini, ada banyak peluang untuk penelitian lebih lanjut yang dapat memperluas pemahaman kita tentang bagaimana sistem kecerdasan buatan melakukan tugas logika visual, khususnya berkaitan dengan gambar mockup. Pertama, penelitian ini menggunakan dataset gambar mockup statis dua dimensi. Dengan menggunakan prototipe atau mockup interaktif seperti animasi antarmuka pengguna, simulasi navigasi aplikasi, atau mockup dalam lingkungan Augmented Reality (AR) atau Virtual Reality (VR), penelitian lanjutan dapat mengembangkan studi serupa. Menguji kemampuan

AI untuk memahami elemen interaktif dan temporal bukan hanya elemen visual statis adalah penting. Kedua, variasi model AI multimodal dapat menjadi perhatian utama. Selain ChatGPT, Anda dapat menguji dan membandingkan model terbaru seperti Google Gemini, LLaVA (Large Language and Vision Assistant), MiniGPT-4, dan Claude dari Anthropic. Perbandingan ini akan menambah literatur tentang bagaimana berbagai sistem multimodal melakukan tugas reasoning visual. Ketiga, penelitian lebih lanjut harus menggunakan prompt dinamis dan adaptif daripada prompt tetap. Memahami bagaimana sensitivitas model AI terhadap variasi instruksi pengguna dapat mempengaruhi hasil interpretasi gambar sangat penting. Selain itu, penelitian kuantitatif juga dapat dilakukan pada pendekatan berbasis prompt engineering. Keempat, studi ini dapat diperluas dengan melibatkan responden manusia (end-users) untuk menilai seberapa dapat dipercaya, relevan, dan bermanfaat output dari masing-masing sistem AI. Penelitian ini akan menjembatani aspek teknis dengan pengalaman pengguna, terutama bagi pengguna di bidang desain, pendidikan, dan pemasaran digital. Kelima, pengembangan sistem hybrid yang menggabungkan kecepatan AI Blackbox dengan kedalaman reasoning ChatGPT dapat menjadi tujuan penelitian yang inovatif. Penelitian selanjutnya akan memperkuat dasar ilmiah tentang AI multimodal dan membantu menciptakan sistem visual intelligence yang lebih efisien, akurat, dan sesuai kebutuhan industri saat ini. Penelitian ini juga dapat merancang arsitektur pipeline yang secara otomatis memilih sistem berdasarkan tingkat kompleksitas gambar atau kebutuhan pengguna akhir.

DAFTAR PUSTAKA

- [1] D. Durgam, N. Anandhan, and R. Pathak, "AI Image Generation: Emerging Trends and Its Impact on UI/UX Design."
- [2] H. Dave, S. Sonje, J. Pardeshi, S. Chaudhari, and P. Raundale, "A survey on Artificial Intelligence based techniques to convert User Interface design mock-ups to code," in *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 28–33. doi: 10.1109/ICAIS50930.2021.9395994.
- [3] M. Samir, A. Elsayed, and M. I. Marie, "A Model for Automatic Code Generation from High Fidelity Graphical User Interface Mockups using Deep Learning Techniques," 2024. [Online]. Available: www.ijacsa.thesai.org
- [4] M. I. Baig and E. Yadegaridehkordi, "ChatGPT in the higher education: A systematic literature review and research challenges," *Int J Educ Res*, vol. 127, Jan. 2024, doi: 10.1016/j.ijer.2024.102411.
- [5] O. Loyola-Gonzalez, "Black-box vs. White-Box: Understanding their advantages and weaknesses from a practical point of view," 2019, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2019.2949286.
- [6] S. Wang, M. A. Qureshi, L. Miralles-Pechuan, T. Huynh-The, T. R. Gadekallu, and M. Liyanage, "Explainable AI for 6G Use Cases: Technical Aspects and Research Challenges," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 2490–2540, 2024, doi: 10.1109/OJCOMS.2024.3386872.
- [7] H. Li, P. Wang, C. Shen, and A. Van Den Hengel, "Visual Question Answering as Reading Comprehension."
- [8] J. Hussain *et al.*, "Model-based adaptive user interface based on context and user experience evaluation," *Journal on Multimodal User Interfaces*, vol. 12, no. 1, pp. 1–16, Mar. 2018, doi: 10.1007/s12193-018-0258-2.
- [9] X. Xu *et al.*, "A Comprehensive Review on Synergy of Multi-Modal Data and AI Technologies in Medical Diagnosis," Mar. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/bioengineering11030219.
- [10] S. Hinterstoisser, V. Lepetit, P. Wohlhart, and K. Konolige, "On Pre-Trained Image Features and Synthetic Images for Deep Learning."
- [11] R. Ashraf *et al.*, "Deep Convolution Neural Network for Big Data Medical Image Classification," *IEEE Access*, vol. 8, pp. 105659–105670, 2020, doi: 10.1109/ACCESS.2020.2998808.
- [12] S. Jamil, M. Jalil Piran, and O. J. Kwon, "A Comprehensive Survey of Transformers for Computer Vision," May 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/drones7050287.
- [13] "Multimodal artificial intelligence in industry: Integrating text, image, and audio for enhanced applications across sectors 1 Dimple Patil."
- [14] R. Javan, T. Kim, and N. Mostaghni, "GPT-4 Vision: Multi-Modal Evolution of ChatGPT and Potential Role in Radiology," *Cureus*, Aug. 2024, doi: 10.7759/cureus.68298.
- [15] O. V. Johnson, O. Mohammed Alyasiri, D. Akhtom, and O. E. Johnson, "Image Analysis through the lens of ChatGPT-4," *Journal of Applied Artificial Intelligence*, vol. 4, no. 2, pp. 31–46, Dec. 2023, doi: 10.48185/jaai.v4i2.870.
- [16] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, Sep. 2018, doi: 10.1109/ACCESS.2018.2870052.
- [17] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput Surv*, vol. 51, no. 5, Sep. 2018, doi: 10.1145/3236009.
- [18] V. Hassija *et al.*, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," Jan. 01, 2024, *Springer*. doi: 10.1007/s12559-023-10179-8.
- [19] E. Akça and Ö. Ö. Tanrıöver, "Exploring Multimodal Large Language Models ChatGPT-4 and Bard for Visual Complexity Evaluation of Mobile User Interfaces," *Traitement du Signal*, vol. 41, no. 5, pp. 2673–2681, Oct. 2024, doi: 10.18280/ts.410540.
- [20] S. Bresciani and M. J. Eppler, "The pitfalls of visual representations: A review and classification of common errors made while designing and interpreting visualizations," *Sage Open*, vol. 5, no. 4, Dec. 2015, doi: 10.1177/2158244015611451.