



Pengembangan Sistem Penilaian Otomatis *User Interface Website* Institusi Berbasis AI Menggunakan Metode *Expert Review* dan YOLOv11

Tsabitha Putri Anjani¹, Yudistira Bagus Pratama², Rike Pradila³

^{1,2,3} Universitas Muhammadiyah Bangka Belitung

Email: tsabitha266@gmail.com

Article Info

Article history:

Received April 15, 2026

Revised April 23, 2026

Accepted Mei 01, 2026

Keywords:

Artificial Intelligence, Heuristic Evaluation, User Interface, Institutional Website, YOLOv11.

ABSTRACT

The visual quality of institutional website user interfaces (UI) plays a vital role in building credibility, yet conventional manual evaluation methods are often constrained by evaluator subjectivity and time inefficiency. Therefore, this study aims to develop an automated assessment system that integrates Computer Vision and Large Language Model (LLM) technologies. This integration is designed to provide a more objective, rapid, and measurable design audit compared to traditional approaches. Using the Waterfall system development method, this research trained the YOLOv11 architecture model on a dataset comprising 100 screenshots of institutional websites in the Bangka Belitung Islands, annotated via the Roboflow platform. The training process was conducted for 300 epochs to specifically detect ten micro UI elements. The detected elements were then qualitatively analyzed by Google Gemini 2.5 Flash to generate assessments based on expert heuristic principles. Technical evaluation results show that the YOLOv11 model achieved stable detection performance with $mAP@50$ values ranging from 0.6 to 0.7. User Acceptance Testing (UAT), involving participants with backgrounds as developers, UI designers, and general users, yielded a final feasibility score of 88.42% with a "Very Good" rating. These findings indicate that AI integration effectively automates the audit process and enhances work efficiency, although label visualization on dense elements requires further optimization.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Article Info

Article history:

Received April 15, 2026

Revised April 23, 2026

Accepted Mei 01, 2026

Keywords:

Artificial Intelligence, Heuristic Evaluation, User Interface, Website Institusi, YOLOv11.

ABSTRACT

Kualitas visual antarmuka pengguna (UI) pada *website* institusi memegang peranan vital dalam membangun kredibilitas, namun metode evaluasi manual seringkali terkendala oleh subjektivitas penilai dan inefisiensi waktu. Oleh karena itu, penelitian ini bertujuan mengembangkan sistem penilaian otomatis yang mengintegrasikan teknologi *Computer Vision* dan *Large Language Model* (LLM). Integrasi ini dirancang untuk memberikan audit desain antarmuka yang jauh lebih objektif, cepat, dan terukur dibandingkan pendekatan konvensional. Menggunakan metode pengembangan sistem *Waterfall*, penelitian ini melatih model arsitektur YOLOv11 pada dataset yang terdiri dari 100 citra tangkapan layar *website* institusi di Kepulauan Bangka Belitung yang dianotasi melalui platform Roboflow. Proses pelatihan dilakukan selama 300 *epoch* untuk mendeteksi sepuluh elemen UI mikro secara spesifik. Elemen yang terdeteksi kemudian dianalisis secara kualitatif oleh kecerdasan buatan Google Gemini 2.5 Flash untuk menghasilkan penilaian berdasarkan prinsip heuristik pakar. Hasil evaluasi teknis menunjukkan model YOLOv11 mencapai performa deteksi yang sangat stabil dengan nilai $mAP@50$ berkisar antara 0,6 hingga 0,7. Pengujian *User Acceptance Testing* (UAT) yang



melibatkan partisipan dengan latar belakang pengembang, desainer UI, dan pengguna umum menghasilkan skor kelayakan akhir sebesar 88,42% dengan predikat "Sangat Baik". Temuan ini mengindikasikan bahwa integrasi AI efektif dalam mengotomatisasi proses audit dan meningkatkan efisiensi kerja, meskipun visualisasi label pada elemen padat masih memerlukan optimasi lebih lanjut.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Tsabitha Putri Anjani
Universitas Muhammadiyah Bangka Belitung
Email: tsabitha266@gmail.com

PENDAHULUAN

User Interface (UI) merupakan komponen penting dalam sistem digital, terutama pada *website* institusi yang berfungsi menyampaikan informasi publik dan layanan akademik. UI yang dirancang dengan baik tidak hanya menjadi penghubung interaksi, tetapi juga memengaruhi kepuasan, persepsi kredibilitas, serta efektivitas pengguna dalam mengakses informasi. Menurut Lněnička et al. (2024), kualitas desain UI berperan penting dalam meminimalkan kebingungan pengguna dan meningkatkan retensi pengunjung.

Prinsip visual seperti keterbacaan, konsistensi *layout*, keseimbangan komposisi, serta kejelasan navigasi menjadi faktor fundamental dalam desain antarmuka yang efektif (Wu et al., 2023). Sejalan dengan pentingnya prinsip tersebut, hasil observasi terhadap *website* institusi di Kepulauan Bangka Belitung menunjukkan masih adanya permasalahan pada penerapan desain antarmuka. Kondisi seperti tata letak elemen yang tidak proporsional dan kontras warna yang lemah menunjukkan bahwa penerapan prinsip UI modern di tingkat lokal belum optimal.

Padahal, peran *website* institusi sangat krusial dalam menunjang transparansi dan akses layanan publik. Kualitas visual yang baik akan langsung berdampak pada kepercayaan masyarakat terhadap institusi tersebut. Oleh karena itu, diperlukan suatu sistem evaluasi UI yang lebih objektif dan efisien agar kualitas tampilan *website* institusi dapat terukur secara berkelanjutan.

Evaluasi kualitas UI selama ini umumnya dilakukan melalui pendekatan manual seperti *heuristic evaluation*, *usability testing*, dan *focus group discussion*. Pendekatan ini sangat bergantung pada persepsi subjektif evaluator, sehingga hasilnya rentan terhadap bias dan memerlukan biaya besar (Kang et al., 2024). Selain itu, metode manual kurang skalabel untuk institusi yang memiliki banyak variasi halaman yang terus diperbarui.

Sebagai solusi atas keterbatasan evaluator manusia, teknologi *Large Language Model* (LLM) seperti Gemini menawarkan kemampuan penalaran multimodal yang canggih untuk difungsikan sebagai *expert validator* otomatis. Integrasi ini memungkinkan sistem untuk tidak hanya mendeteksi keberadaan elemen, tetapi juga mengevaluasi kesesuaiannya terhadap standar *usability* secara naratif (Lu et al., 2024). Proses evaluasi ini menjembatani analisis mesin yang kaku dengan interpretasi desain manusia melalui kecepatan yang jauh lebih tinggi.



Seiring dengan perkembangan teknologi *Computer Vision*, metode *object detection* menjadi fondasi utama untuk mengenali struktur antarmuka. Model YOLOv11 dipilih sebagai algoritma terbaru dengan keunggulan akurasi tinggi dan efisiensi waktu dalam mendeteksi elemen visual mikro (Ali dan Zhang, 2024). *Framework* Ultralytics dan platform Roboflow mendukung pelatihan model ini untuk menghasilkan anotasi elemen yang presisi (Tian et al., 2025).

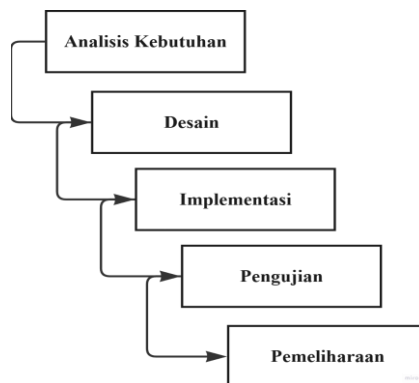
Berdasarkan penjelasan di atas, penelitian ini berfokus pada pengembangan sistem penilaian otomatis antarmuka *website* institusi berbasis AI. Sistem ini mengintegrasikan YOLOv11 untuk deteksi visual dan Gemini sebagai *expert validator* kualitatif. Diharapkan, hasil penelitian ini memberikan manfaat praktis berupa peningkatan efisiensi, objektivitas, dan konsistensi dalam penilaian desain antarmuka *website* institusi.

METODE PENELITIAN

Penelitian ini menggunakan metode kualitatif dengan jenis eksperimen terapan yang berfokus pada pengembangan sistem penilaian otomatis antarmuka *website* berbasis *Artificial Intelligence* (AI). Pendekatan kualitatif dipilih karena hasil keluaran model deteksi membutuhkan interpretasi pakar melalui proses *expert review* untuk memvalidasi relevansinya (Abubakar, 2021). Kejelasan waktu pelaksanaan dan rancangan metodologis menjadi dasar penting dalam menjaga konsistensi antara tahap perancangan, implementasi, serta pelaporan hasil (Kustiawan et al., 2025).

Penelitian dilakukan pada *website* institusi pemerintahan dan pendidikan di Provinsi Kepulauan Bangka Belitung yang masih menunjukkan keterbatasan kualitas antarmuka (Ronaldo et al., 2024). Sebanyak 100 tangkapan layar (*screenshot*) halaman utama dari berbagai situs resmi tersebut dikumpulkan secara acak untuk dijadikan *dataset* visual. Pengambilan data ini dilakukan secara manual melalui akses publik sesuai dengan etika penelitian digital tanpa melakukan peretasan (Priyanda et al., 2022).

Pada tahap pengembangan, penelitian ini menggunakan metode *waterfall* karena alurnya yang sistematis sangat sesuai untuk proyek yang menuntut dokumentasi terstruktur (Saravanas dan Curinga, 2023). Pendekatan ini dipilih karena spesifikasi kebutuhan sistem dinilai sudah stabil dan tidak akan mengalami perubahan fundamental di tengah proses (Thesing et al., 2021). Selain itu, penerapan metode *waterfall* sebelumnya telah terbukti efektif untuk pengembangan sistem informasi berbasis *website* di tingkat daerah (Ronaldo et al., 2024).



Gambar 1. Alur Tahapan Metode Waterfall
(Sumber: Asmarajaya et al., 2021)

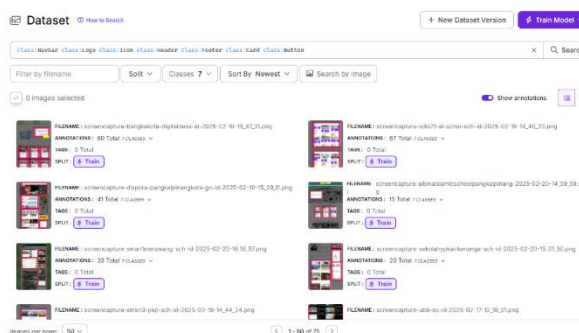
Tahapan metode *waterfall* dalam penelitian ini mencakup lima fase berurutan, yaitu analisis kebutuhan, desain, implementasi, pengujian, dan pemeliharaan (Alam et al., 2022). Fase analisis hingga implementasi berfokus pada penyiapan *dataset* menggunakan platform Roboflow dan pelatihan model deteksi (Krisnha et al., 2025). Sementara itu, fase pengujian memastikan sistem sesuai kriteria dan fase pemeliharaan merancang strategi perbaikan terhadap anomali deteksi di masa depan (Yas et al., 2023).

HASIL DAN PEMBAHASAN

Analisis Kebutuhan

Penelitian ini diawali dengan penentuan model deteksi objek YOLOv11 yang dinilai paling sesuai untuk mengenali elemen antarmuka pengguna pada *website* institusi. Pemilihan model ini didasarkan pada arsitekturnya yang terbukti sangat stabil dalam mendeteksi berbagai objek visual berukuran kecil maupun rapat. Selain itu, kemampuan inferensi YOLOv11 yang cepat mendukung kebutuhan sistem evaluasi otomatis agar dapat beroperasi secara efisien bagi pengguna.

Tahap selanjutnya berfokus pada proses identifikasi elemen antarmuka yang akan menjadi target deteksi utama oleh sistem. Peneliti menetapkan sepuluh kategori yang paling memengaruhi kualitas tampilan visual, yaitu *header*, *logo*, *navbar*, *icon*, *buttons*, *footer*, *card*, *scale*, *font*, dan *colour*. Pemilihan kesepuluh kategori ini dirancang agar model AI mampu mempelajari serta mengenali komponen yang umum digunakan pada desain *website* institusi.



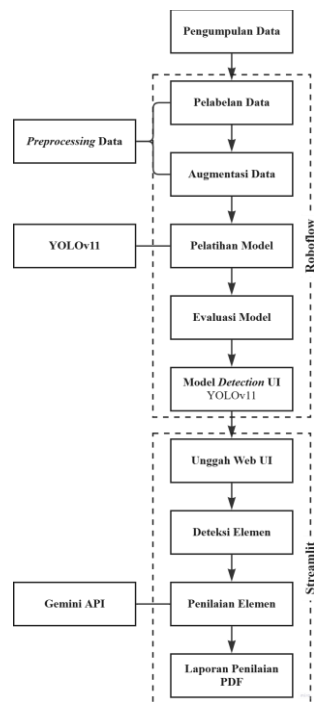
Gambar 2. Contoh Dataset Web Istitusi



Proses pengumpulan data dilakukan dengan menghimpun 100 tangkapan layar halaman depan dari berbagai *website* institusi di wilayah Kepulauan Bangka Belitung. Jumlah ini dianggap sudah memadai untuk membangun model deteksi awal karena adanya kesamaan pola struktur tampilan pada situs-situs tersebut. Skala data ini memungkinkan proses anotasi dan pelatihan model dilakukan secara efektif tanpa menghilangkan keragaman visual yang dibutuhkan.

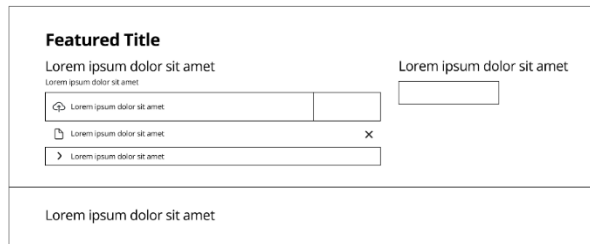
Desain

Tahapan perancangan sistem mencakup pembuatan arsitektur alur kerja deteksi antarmuka yang mengandalkan model YOLOv11 sebagai pemroses visual utama. Model deteksi tersebut kemudian diintegrasikan dengan mekanisme penilaian heuristik berbasis AI melalui penyusunan *wireframe* aplikasi antarmuka Streamlit. Pendekatan desain komprehensif ini bertujuan menghasilkan sistem evaluasi antarmuka yang akurat secara teknis, mudah dioperasikan, dan mendukung proses audit otomatis.



Gambar 3. Arsitektur Sistem

Arsitektur sistem yang dikembangkan oleh peneliti memuat lima alur kerja utama yang saling terintegrasi satu sama lain. Alur tersebut meliputi tahapan pengumpulan tangkapan layar, pelabelan data di Roboflow, pelatihan model YOLOv11, hingga pelaksanaan proses deteksi otomatis. Setelah elemen berhasil dideteksi, hasilnya langsung diteruskan menuju *Large Language Model* Gemini AI untuk dianalisis secara kualitatif dan diekspor menjadi laporan PDF.



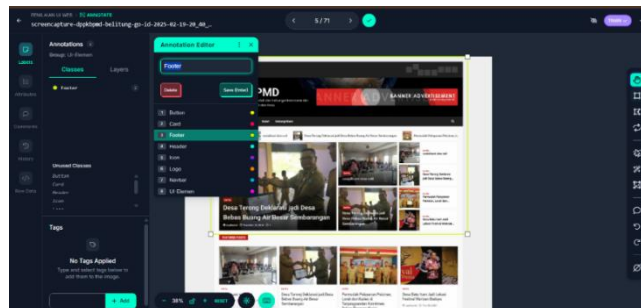
Gambar 4. Wireframe Sistem Penilaian

Rancangan *wireframe* pada antarmuka aplikasi Streamlit dibagi menjadi tiga bagian fungsional utama untuk memaksimalkan pengalaman pengguna. Ketiga bagian tersebut adalah area untuk mengunggah *screenshot*, panel untuk menampilkan hasil deteksi beserta penilaian AI, dan fasilitas unduh laporan. Desain tata letak fitur ini dibuat secara berurutan dan minimalis agar pengguna dapat memahami alur analisis antarmuka dengan sangat mudah.

Implementasi

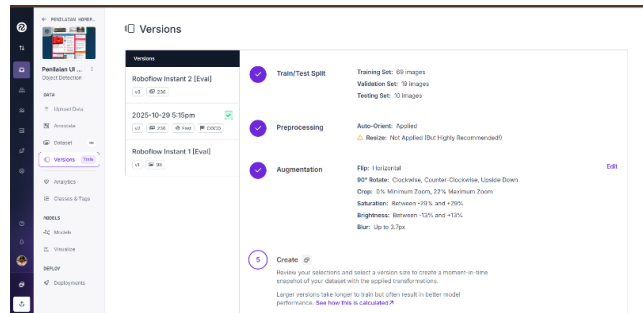
a. Preprocessing Data

Tahap pra-pemrosesan data bertujuan kuat untuk memastikan kualitas maupun konsistensi dari seluruh tangkapan layar *website* institusi yang telah dikumpulkan. Kualitas gambar yang seragam akan sangat membantu model dalam mendeteksi elemen UI dengan tingkat akurasi yang lebih tinggi. Oleh karena itu, seluruh gambar disortir ketat untuk memastikan resolusinya memadai serta terbebas dari distorsi yang dapat mengganggu performa deteksi.



Gambar 5. Pelabelan Elemen di Roboflow

Langkah berikutnya adalah mengeksekusi anotasi secara manual pada setiap elemen UI menggunakan bantuan *platform* Roboflow. Setiap gambar di dalam dataset diberi penanda *bounding box* yang presisi sesuai dengan sepuluh kategori elemen antarmuka. Proses pelabelan manual ini dilakukan secara teliti untuk menjamin akurasi kelas objek sekaligus memberikan contoh visual representatif bagi model AI.

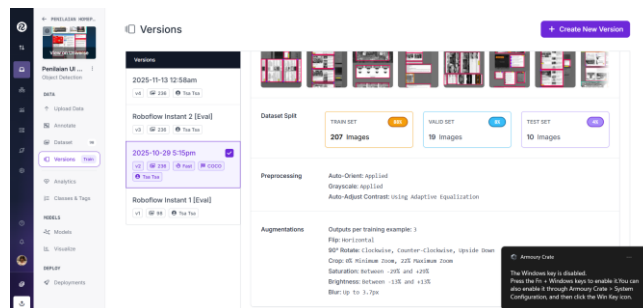


Gambar 6. Proses Augmentasi

Setelah proses anotasi rampung secara keseluruhan, dataset tersebut langsung diolah lebih lanjut melalui tahap augmentasi. Transformasi yang diterapkan mencakup rotasi ringan, *horizontal flip*, penyesuaian kecerahan, hingga penambahan efek *blur* untuk meningkatkan variasi data dan mencegah *overfitting*. Dataset yang telah diperkaya variasi visualnya ini kemudian diekspor dalam format standar agar kompatibel dengan *pipeline* pelatihan YOLOv11.

b. Training Data YOLOv11

Pelatihan model deteksi dilakukan sepenuhnya di dalam *platform* Roboflow tanpa mengandalkan lingkungan komputasi dari pihak eksternal. Proses ini memanfaatkan fitur augmentasi otomatis yang secara cerdas mampu meningkatkan keragaman data tanpa merusak karakteristik utama elemen. Melalui penambahan augmentasi orientasi, model dilatih agar lebih tangguh dalam mengenali elemen desain meskipun gambar *input* mengalami perubahan sudut.



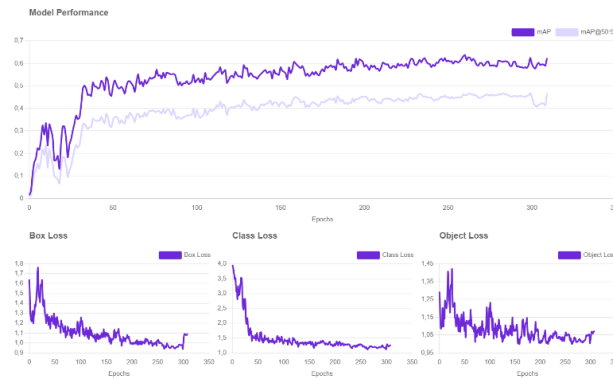
Gambar 7. Hasil Pelatihan Model

Selain penyesuaian orientasi objek, Roboflow juga menerapkan augmentasi skala visual yang dipadukan dengan manipulasi tingkat saturasi warna. Penyesuaian kecerahan dan kontras turut diaplikasikan untuk menyimulasikan perbedaan kondisi desain antarmuka yang rentan tidak seragam antar *website*. Setelah seluruh variasi ini terpasang secara otomatis, sistem melanjutkan proses pemisahan dataset tersebut ke dalam beberapa *subset* untuk evaluasi.

Dataset diekstraksi ke dalam tiga *subset* utama yang mencakup 207 citra latih, 19 citra validasi, dan 10 citra uji murni. Strategi penerapan augmentasi memang dikhususkan hanya pada *Training Set* untuk mencegah terjadinya fenomena kebocoran data. Keputusan ini menjamin bahwa metrik evaluasi mAP yang dihasilkan nantinya benar-benar valid karena diuji dengan data baru.

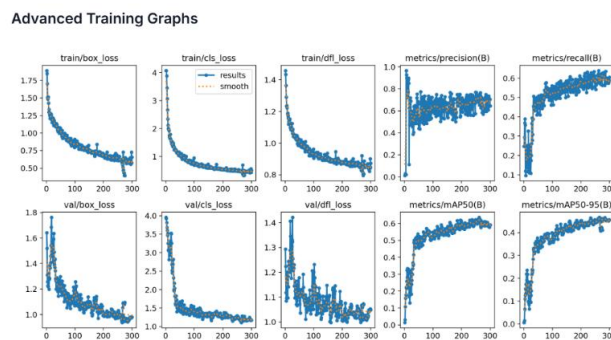
c. Data Evaluation YOLOv11

Kinerja model YOLOv11 dievaluasi secara mendalam dengan menganalisis parameter statistik utama untuk memastikan stabilitas deteksinya. Analisis ini melibatkan pemantauan terhadap metrik *loss*, *Mean Average Precision* (mAP), serta perhitungan nilai presisi dan *recall* secara keseluruhan. Berdasarkan grafik pelatihan, indikator *Box Loss*, *Class Loss*, dan *Object Loss* secara konsisten menunjukkan tren penurunan yang signifikan seiring bertambahnya *epoch*.



Gambar 8. Data Grafik dan Metrik mAP

Pola kurva *loss* yang melandai pada akhir masa pelatihan memberikan konfirmasi bahwa model berhasil mencapai kondisi konvergen. Pergerakan grafik ini membuktikan tidak adanya sedikit pun indikasi ketidakstabilan maupun *overfitting* selama proses pembelajaran berlangsung. Pada evaluasi akurasi, nilai mAP@0.5 terpantau mengalami peningkatan bertahap hingga stabil pada kisaran 0.6–0.7 yang menunjukkan performa deteksi solid.



Gambar 9. Metrik Precision dan Recall

Pencapaian nilai presisi yang mendekati angka 0.8 mengindikasikan kemampuan model dalam membedakan setiap elemen visual dengan sangat baik. Kondisi ini membuat tingkat kesalahan deteksi palsu atau *false positive* dapat ditekan menjadi relatif sangat rendah. Keseimbangan antara presisi dan *recall* ini menjamin bahwa sistem akurat dan tidak akan melewatkan komponen antarmuka esensial untuk dinilai.



d. Integrasi Model Penilaian Menggunakan Gemini

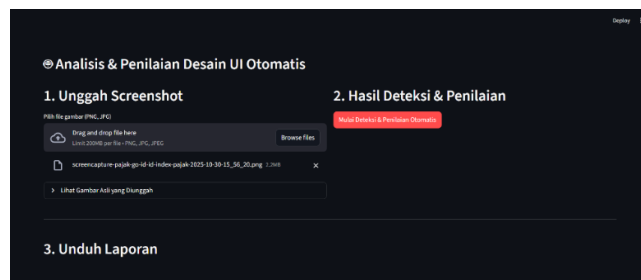
Pasca pelatihan, model deteksi objek YOLOv11 langsung diintegrasikan dengan modul penalaran multimodal Gemini 2.5 Flash. Integrasi ini dibangun menggunakan arsitektur antarmuka Streamlit berbasis Python untuk menjembatani data visual dan analisis kualitatif. Peneliti juga mengembangkan algoritma pemetaan khusus yang secara otomatis memberikan identitas unik pada setiap objek agar model bahasa memahami konteks spasial.

Pemberian identitas unik ini berperan penting agar Gemini dapat memberikan komentar yang terarah pada komponen antarmuka tertentu. Pendekatan ini mencegah AI mengeluarkan analisis normatif yang terlalu umum terhadap keseluruhan tangkapan layar *website*. *Prompt* juga disusun secara ketat dengan memberikan citra visual beserta koordinat agar model selalu menghasilkan penilaian berformat JSON.

Penggunaan model Gemini-2.5-flash ini secara khusus dipilih karena kemampuannya yang sangat unggul dalam hal kecepatan inferensi. Model ini juga mendukung penalaran multimodal yang jauh lebih komprehensif untuk menangani analisis *User Interface*. Sebagai penyempurna, kerangka kerja ini dihubungkan dengan pustaka ReportLab untuk menghasilkan dokumen evaluasi format PDF secara otomatis.

e. Deployment Menggunakan Streamlit

Fase *deployment* sistem dieksekusi melalui *framework* Streamlit karena terbukti andal dalam memfasilitasi pembangunan aplikasi web berbasis Python. Pemilihan alat ini memungkinkan terciptanya aplikasi yang responsif dan interaktif tanpa menuntut konfigurasi *frontend* rumit. Sistem ini juga menghadirkan integrasi mulus antara deteksi *real-time* YOLOv11 dengan layanan evaluasi AI untuk meringankan beban komputasi pengguna.



Gambar 10. Interface Streamlit

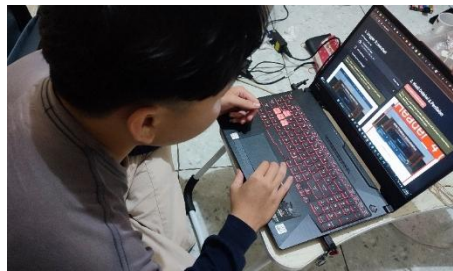
Melalui antarmuka yang intuitif ini, pengguna hanya diminta untuk mengunggah *screenshot website* institusi ke dalam panel aplikasi. Sistem kemudian akan mengirimkan citra tersebut ke *server* secara otomatis untuk diproses oleh algoritma deteksi objek. Hasil deteksi berupa *bounding box* langsung diteruskan ke model Gemini untuk dievaluasi kualitatif dan ditampilkan di layar tanpa langkah manual tambahan.

Pengujian Sistem

a. Blackbox Testing

Kelayakan operasional sistem diuji menggunakan metode *Blackbox Testing* sebelum melibatkan partisipasi pengguna akhir. Pengujian ini menggunakan pendekatan *State Transition* yang berfokus penuh pada validasi kelancaran perpindahan status di dalam aplikasi.

Tujuannya adalah memastikan seluruh alur kerja sistem dapat tereksekusi secara sempurna sesuai dengan rancangan logika awal.



Gambar 11. Proses Pengujian Sistem

Eksekusi skenario pengujian dilakukan secara menyeluruh baik pada lingkungan *deployment* lokal maupun *cloud*. Langkah ini berguna untuk memverifikasi bahwa respons interaktif sistem tidak mengalami degradasi kualitas di berbagai kondisi operasional. Hasil pengujian membuktikan bahwa sistem menunjukkan stabilitas fungsional yang sangat andal pada kelima skenario transisi utamanya.

Tabel 1. Hasil *Blackbox Testing* (Test Result)

Transisi	Skenario Pengujian	Ekspektasi	Hasil Aktual	Keterangan
T1	User mengunggah tangkapan layar <i>website</i> institusi berformat.PNG	Berpindah ke status "Uploaded" dan menampilkan gambar asli dengan rasio yang benar.	Gambar berhasil diunggah, pratinjau muncul dalam < 2 detik tanpa distorsi.	Berhasil
T2	User memulai proses deteksi pada gambar yang telah diunggah	Berpindah halaman menampilkan gambar hasil anotasi (kotak deteksi) pada elemen UI.	Halaman berpindah, elemen seperti <i>Navbar</i> , <i>Button</i> , dan Logo terkotak dengan label yang sesuai.	Berhasil
T3	User melihat detail analisis heuristik dari AI	Menampilkan narasi kritik desain yang relevan dengan elemen yang terdeteksi.	Narasi muncul dalam format terstruktur (point-point), dapat dibaca dengan jelas.	Berhasil
T4	User mengunduh hasil audit dalam bentuk laporan	File PDF terunduh berisi gabungan gambar visual dan teks analisis.	File "Laporan_Evaluasi.pdf" berhasil diunduh dan dapat dibuka dengan sempurna.	Berhasil
T5	User mereset aplikasi untuk menilai gambar lain	Kembali ke halaman awal, data lama hilang dari memori.	Halaman kembali bersih (<i>Home</i>), siap menerima input baru.	Berhasil

Seluruh fungsi logika sistem dinyatakan beroperasi dengan sangat baik meskipun terdapat catatan kecil pada aspek visualisasinya. Label deteksi pada area navigasi yang sangat padat cenderung saling bertumpuk ketika ditampilkan di layar aplikasi. Namun demikian, hal ini tidak menggugurkan status keberhasilan sistem karena fungsionalitas deteksinya tetap akurat dan siap diujikan kepada pengguna.



b. User Acceptance Testing

Tahap validasi subjektivitas dari perspektif audiens dilakukan melalui skema *User Acceptance Testing* (UAT) untuk mengukur tingkat penerimaan pengguna, mengevaluasi aspek kemudahan penggunaan, serta membuktikan hipotesis efisiensi waktu kerja. Pengujian komprehensif ini melibatkan 20 responden heterogen yang dipilih secara *purposive*, yang terdiri dari praktisi desainer antarmuka, mahasiswa berlatar belakang teknologi informasi, dan beberapa pengguna awam. Data kuesioner yang terkumpul dari rentang Skala Likert kemudian dikonversi menjadi persentase kelayakan dengan menggunakan rumus standar sistem informasi agar interpretasinya lebih mudah dipahami secara analitis oleh para peneliti.

Berdasarkan kalkulasi kuantitatif, variabel fungsionalitas mencatatkan rata-rata sebesar 87,83% yang mengonfirmasi bahwa ketepatan deteksi elemen dan keandalan fitur sistem telah memenuhi ekspektasi pengguna secara meyakinkan. Poin tertinggi berhasil diraih pada metrik ketepatan klasifikasi label yang memvalidasi efektivitas pelatihan YOLOv11, meskipun sempat terjadi sedikit penurunan skor pada variasi tata letak yang terlalu ekstrem akibat tantangan generalisasi AI. Kinerja sistem juga direspons positif dengan skor rata-rata mencapai 87,25%, yang menempatkan keandalan fungsional aplikasi ke dalam predikat Sangat Baik.

Stabilitas aplikasi sempat menjadi item dengan skor terendah karena tingginya beban pemrosesan inferensi berisiko menimbulkan jeda singkat pada koneksi internet lambat, namun kelancaran perpindahan navigasi antar menu dinilai sangat responsif oleh para pengguna. Variabel pengalaman antarmuka sukses meraih skor sebesar 88,78%, di mana mayoritas responden sangat mengapresiasi tingkat keterbacaan teks label hasil deteksi serta tata letak visual aplikasi yang tersusun secara logis. Kemudahan interaksi ini membuktikan bahwa arsitektur antarmuka UX yang diimplementasikan telah berjalan optimal, sehingga aplikasi dapat digunakan dengan lancar oleh siapa saja meskipun tanpa panduan operasional khusus.

Variabel efisiensi berhasil meraih skor rata-rata tertinggi pada angka 89,83%, yang secara langsung memvalidasi kemampuan sistem ini dalam memangkas waktu kerja jika dibandingkan dengan proses *review* konvensional. Fitur otomasi penyusunan laporan berbasis AI ini terbukti sangat diapresiasi oleh pengguna karena sukses menghilangkan hambatan kerja administratif yang selama ini dinilai sangat melelahkan pada evaluasi manual. Jika diakumulasikan secara keseluruhan dari keempat variabel pengujian, sistem penilaian otomatis antarmuka pengguna berbasis kecerdasan buatan ini berhasil mengamankan skor akhir kelayakan sebesar 87,75% dari seluruh responden yang terlibat.

Tabel 2. Hasil Akhir Perhitungan UAT

No	Variabel	Nilai Bobot (%)	Keterangan
1	Fungsionalitas Sistem	87%	Sangat Baik
2	Kinerja Sistem	87%	Sangat Baik
3	Pengalaman dan Tampilan Antarmuka	88%	Sangat Baik
4	Efisiensi dan Produktivitas	89%	Sangat Baik

Capaian persentase kelayakan tersebut secara mantap menempatkan sistem ini ke dalam kategori “Sangat Baik” menurut standar evaluasi perangkat lunak. Predikat ini secara langsung menjadi justifikasi kuat atas kesiapan fungsional aplikasi untuk diimplementasikan secara luas.



Sistem ini dapat segera dimanfaatkan sebagai instrumen pengukur standar dalam mendukung proses audit desain *website* di berbagai institusi.

Pemeliharaan

Fase pemeliharaan sistem dilaksanakan sebagai langkah proaktif untuk mengevaluasi dan merumuskan strategi perbaikan perangkat lunak di masa depan. Pendekatan ini berguna untuk memantau berbagai anomali teknis yang mungkin muncul ketika aplikasi dihadapkan pada data *website* baru. Analisis mendalam menunjukkan bahwa sistem terkadang masih mengalami kebingungan klasifikasi akibat kemiripan bentuk visual, seperti ikon navigasi yang terdeteksi sebagai logo.

Selain masalah klasifikasi elemen grafis yang ambigu, kecerdasan buatan juga sering terkecoh oleh desain blok warna dekoratif pada *website* modern. Kondisi tersebut menyebabkan model keliru melabeli area latar belakang kosong sebagai komponen struktural seperti *header* maupun *footer*. Untuk mengatasinya, direkomendasikan agar pengembangan selanjutnya lebih berfokus pada pengayaan variasi data latih YOLOv11 dan penyesuaian algoritma perenderan label antarmuka.

KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan sistem penilaian otomatis yang menggabungkan kemampuan deteksi visual *Deep Learning* (YOLOv11) dengan kemampuan penalaran semantik *Large Language Model* (Gemini 2.5 Flash). Integrasi ini terbukti mampu mereplikasi proses *Heuristic Evaluation* manual, di mana YOLOv11 bertindak sebagai "mata" yang menginventarisasi elemen UI (seperti *Navbar*, *Card*, *Footer*), dan Gemini bertindak sebagai "otak" yang memberikan penilaian kualitatif berdasarkan prinsip desain. Evaluasi teknis terhadap model YOLOv11 menunjukkan performa yang tangguh dalam mengenali elemen antarmuka *website* institusi. Grafik pelatihan menunjukkan konvergensi *loss* (*Box*, *Class*, *Object*) yang stabil dan tren peningkatan nilai *Mean Average Precision* (mAP) yang konsisten hingga akhir *epoch*. Hal ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik untuk mendeteksi berbagai komponen visual *website* dengan tingkat presisi dan *recall* yang seimbang, meminimalkan kesalahan deteksi pada elemen-elemen standar.

Saran

Meskipun sistem telah berfungsi dengan baik dan memenuhi tujuan penelitian, terdapat beberapa keterbatasan yang ditemukan selama fase pengujian disarankan menambah dan memperkaya Data Latihan (*Dataset*). Selain mengatasi masalah pada gaya desain yang tidak umum, penelitian selanjutnya sangat disarankan untuk memperbanyak jumlah contoh gambar yang digunakan untuk melatih sistem. Semakin banyak variasi data gambar yang dipelajari oleh sistem deteksi (YOLOv11), sistem akan semakin tangguh dan pintar dalam mengenali berbagai bentuk elemen rumit, sehingga tidak salah lagi dalam membedakan antara ikon dekoratif atau logo institusi.

**DAFTAR PUSTAKA**

- Abubakar, R. (2021). Pengantar metodologi penelitian. SUKA-Press UIN Sunan Kalijaga. <https://books.google.com/books?id=5ijKEAAAQBAJ>
- Alam, I., Sarwar, N., & Noreen, I. (2022). *Statistical analysis of software development models by six-pointed star framework*. *PLoS One*, 17(4), e0264420. <https://doi.org/10.1371/journal.pone.0264420>
- Ali, M. L., dan Zhang, Z. (2024). *The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection*. *Computers*, 13(12), 336. <https://doi.org/10.3390/computers13120336>
- Asmarajaya, I. K. A., Sanjaya, K. O., Putra, D. M. D. U., Mahendra, G. S., dan Hasanah, F. N. U. (2021). Sistem Informasi Keuangan pada Perusahaan Kost Elit dengan Metode Waterfall. *Swabumi*, 9(2), 100–108. <https://doi.org/10.31294/swabumi.v9i2.10970>
- Kang, J., Kim, J., Yang, M., Park, C., Kim, T., Song, H., dan Han, J. (2024). *Developing an AI-based explainable expert support system for art therapy*. *ACM Transactions on Interactive Intelligent Systems*, 14(4), 30. <https://doi.org/10.1145/3689649>
- Krisnha, A. V., Ramdhan, N. A., dan Premana, A. (2025). Implementasi metode waterfall dalam pengembangan sistem pengajuan produk hukum daerah berbasis web dengan FIFO dan tracking. *Elkom: Jurnal Elektronika dan Komputer*, 18(1), 389–402. <https://doi.org/10.51903/elkom.v18i1.2954>
- Kustiawan, W., Daffa Raihan, M., dan Ibnu Thariq, M. (2025). Pendekatan metode penelitian lokasi dan waktu informan penelitian dan penyusunan. *Jurnal Pendidikan Tambusai*, 9(1), 5051–5055. <https://jptam.org/index.php/jptam/article/view/28023>
- Lněnička, M., Nikiforova, A., Wang, D., dan Bernadini, F. (2024). *UX competitive analysis of smart city open data portals: Usability framework, design recommendations, and a roadmap for sustainable data ecosystems*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4960428>
- Lu, Y., Yang, Y., Zhao, Q., Zhang, C., dan Li, T. J.-J. (2024). *AI assistance for UX: A literature review through human-centered AI*. *arXiv preprint arXiv:2402.06089*. <https://arxiv.org/abs/2402.06089>
- Priyanda, R., Agustina, T. S., Ariantini, N. S., Rusmayani, N. G. A. L., Aslindar, D. A., Ningsih, K. P., Wulandari, S., Putranto, P., Yuniati, I., et al. (2022). Metodologi penelitian kuantitatif. Pradina Pustaka. <https://books.google.co.id/books?id=B5t1EAAAQBAJ>
- Ronaldo, M. D., Putri, C. A., Zaki, M., Firmansyah, D., Wicaksana, S. S., Fauzan, M., Budiarjo, A., dan Tou, N. (2024). *Design and development of Bangka District regional library*



- information system based on website. Bangka Information Technology Journal*, 1(1), 41–49. <https://doi.org/10.33019/5na9mp47>
- Saravanos, A., & Curinga, M. X. (2023). *Simulating the software development lifecycle: The waterfall model. Applied System Innovation*, 6(6), 108. <https://doi.org/10.3390/asi6060108>
- Thesing, T., Feldmann, C., & Burchardt, M. (2021). *Agile versus waterfall project management: Decision model for selecting the appropriate approach to a project. Procedia Computer Science*, 181, 746–756. <https://doi.org/10.1016/j.procs.2021.01.227>
- Tian, Z., Yang, F., Yang, L., Wu, Y., Chen, J., dan Qian, P. (2025). *An optimized YOLOv11 framework for the efficient multi-category defect detection of concrete surface. Sensors*, 25(5), 1291. <https://doi.org/10.3390/s25051291>
- Wu, J., Peng, Y. H., Li, A., Swearngin, A., Bigham, J. P., dan Nichols, J. (2024). *UIClip: A data-driven model for assessing user interface design. arXiv preprint arXiv:2404.12500*. <https://arxiv.org/abs/2404.12500>
- Yas, Q. M., ALazzawi, A., & Rahmatullah, B. (2023). *A comprehensive review of software development life cycle methodologies: Pros, cons, and future directions. Iraqi Journal for Computer Science and Mathematics*, 4(4), Article 14. <https://doi.org/10.52866/ijcsm.2023.04.04.014>