

A Semiotic-Based Usability Framework for Evaluating AI-Generated Visual Communication in Design Education: A UI/UX Testing Approach

Nashiruddin Al-fath¹⁾, Yufiarti²⁾

¹⁾Desain Komunikasi Visual, Fakultas Komunikasi dan Multimedia, Binus University
Kompleks Alam Sutera, Jl. Jalur Sutera Barat Kav. 21, Kota Tangerang, Bante

¹⁾alfath@binus.ac.id

²⁾Psikologi, Fakultas Psikologi, Universitas Negeri Jakarta
Jl. Halimun Raya No.2, Guntur, Kecamatan Setiabudi, Jakarta Selatan 12980

²⁾yufiarti@unj.ac.id

Abstrak

Adopsi platform AI generatif dalam praktik Desain Komunikasi Visual menimbulkan pertanyaan tentang kemampuan visual AI menyampaikan makna komunikatif, bukan sekadar menghasilkan tampilan estetis. Penelitian ini mengembangkan dan memvalidasi *Semiotic UX Evaluation Framework* (SUEF), protokol evaluasi tujuh dimensi yang mengoperasionalkan teori semiotika menjadi metrik terukur dalam *usability testing* berbasis perangkat lunak. Penelitian menggunakan *mixed methods sequential explanatory* dengan melibatkan 32 mahasiswa DKV. Evaluasi dilakukan melalui *System Usability Scale*, *AttrakDiff*, *eye tracking*, *think-aloud protocol*, serta *pipeline* NLP berbasis IndoBERT dan LDA. Hasil menunjukkan visual AI-generated memperoleh skor usabilitas *acceptable* (SUS = 71.4), sedangkan desain manual mencapai kategori *good* (SUS = 81.2). Perbedaan tersebut signifikan secara statistik ($p < .001$, Cohen's $d = 0.85$). *Semiotic scoring* menunjukkan bahwa kesenjangan AI dan desain manual bukan terletak pada aspek estetika. Dimensi *Aesthetic Coherence* hanya berbeda 0.2 poin, sementara kesenjangan terbesar terdapat pada *Cultural Relevance* (3.6 poin) dan *Intentionality Alignment* (3.4 poin). Temuan ini menegaskan bahwa AI generatif telah mendekati paritas estetis, tetapi belum mencapai paritas komunikatif. SUEF terbukti replikatif untuk mengevaluasi kualitas komunikatif konten visual berbasis AI secara sistematis dan dapat mendukung audit desain AI pada konteks pendidikan, penelitian, serta praktik profesional DKV di Indonesia.

Kata kunci: *usability testing, semiotic UX, AI-generated design, komunikasi visual, mixed methods*

Abstract

The adoption of generative AI platforms in Visual Communication Design raises questions about whether AI-generated visuals can convey communicative meaning, not merely produce aesthetic appearances. This study develops and validates the *Semiotic UX Evaluation Framework* (SUEF), a seven-dimensional protocol that operationalizes semiotic theory into measurable metrics within usability testing. The study used a *sequential explanatory mixed methods approach* involving 32 Visual Communication Design students. Evaluation was conducted through the *System Usability Scale*, *AttrakDiff*, *eye tracking*, *think-aloud protocol*, and an NLP pipeline based on IndoBERT and LDA. Results show that AI-generated visuals achieved acceptable usability (SUS = 71.4), while manual designs reached the good category (SUS = 81.2). This difference was statistically significant ($p < .001$, Cohen's $d = 0.85$). *Semiotic scoring* indicates that the gap between AI-generated visuals and manual designs does not lie in aesthetics. *Aesthetic Coherence* differed by only 0.2 points, whereas the largest gaps appeared in *Cultural Relevance* (3.6 points) and *Intentionality Alignment* (3.4 points). These findings confirm that generative AI has approached aesthetic parity but not communicative parity. SUEF is replicable for evaluating AI-based visual content and supporting AI design audits in Visual Communication Design education.

Keywords: *usability testing, semiotic UX, AI-generated design, visual communication, mixed methods*

1. PENDAHULUAN

Perkembangan teknologi generatif *Artificial Intelligence* (Gen-AI) telah mendorong transformasi mendasar dalam praktik Desain Komunikasi Visual (DKV). Platform berbasis model difusi seperti Midjourney, DALL-E 3, dan Adobe Firefly mampu menghasilkan aset visual berkualitas tinggi dari instruksi teks dalam hitungan detik. Kemampuan ini mengubah alur kerja desainer dari proses kreatif manual berbasis keahlian teknis menjadi proses ko-kreatif antara manusia dan sistem algoritmik [1].

Pergeseran ini menghadirkan dua implikasi utama. Di satu sisi, Gen-AI meningkatkan produktivitas eksplorasi visual karena desainer dapat menghasilkan banyak variasi konseptual dalam waktu singkat. Di sisi lain, kemampuan AI dalam memahami dan mereproduksi makna kontekstual, simbolik, dan budaya masih terbatas. Sistem AI bekerja berdasarkan pencocokan pola statistik dari data pelatihan, bukan berdasarkan pemahaman semiotik yang terbentuk melalui pengalaman sosial dan budaya manusia [2, 3].

Dalam konteks Human-Computer Interaction (HCI) dan evaluasi perangkat lunak, pertanyaan penting yang belum banyak dijawab secara empiris adalah apakah visual AI memiliki kualitas usability dan kedalaman makna semiotik yang setara dengan desain manual. Pertanyaan ini penting karena keluaran visual AI tidak cukup dinilai dari aspek estetika, tetapi juga dari kemampuannya menyampaikan pesan kepada audiens tertentu. Semiotika, sebagai ilmu tentang tanda dan pembentukan makna, menyediakan kerangka untuk memahami bagaimana visual menyampaikan pesan dan mengapa pesan yang sama dapat dimaknai secara berbeda oleh pengguna.

Secara teoretis, Ferdinand de Saussure menjelaskan bahwa tanda terdiri atas penanda, yaitu bentuk yang dapat dilihat atau didengar, dan petanda, yaitu konsep yang diwakilinya. Hubungan keduanya bersifat sosial dan kultural, bukan alami [4]. Charles Sanders Peirce kemudian membedakan tanda menjadi ikon, indeks, dan simbol berdasarkan relasinya dengan objek [5]. Pembedaan ini relevan dalam evaluasi visual karena jenis tanda yang digunakan menentukan kemudahan audiens dalam memahami makna, terutama pada konteks budaya yang berbeda.

Meskipun semiotika telah lama digunakan dalam kajian DKV, penerapannya dalam evaluasi UI/UX masih terbatas. Analisis semiotik umumnya dilakukan setelah produk selesai dikembangkan, bersifat kualitatif, dan bergantung pada interpretasi ahli. Hingga kini, belum tersedia protokol terstandar yang mengintegrasikan dimensi semiotika ke dalam usability testing, khususnya untuk mengevaluasi visual yang dihasilkan oleh AI generatif.

Kondisi ini menimbulkan tantangan baru dalam evaluasi visual berbasis AI. Pada desain manual, elemen visual dipilih secara sadar berdasarkan tujuan komunikasi tertentu. Sebaliknya, visual AI dihasilkan melalui pola statistik data pelatihan tanpa pemahaman kontekstual terhadap budaya, audiens, maupun intensi komunikasi. Akibatnya, dapat terjadi semiotic drift, yaitu kesenjangan antara makna yang ingin disampaikan, makna yang direpresentasikan oleh AI, dan makna yang ditafsirkan pengguna [6, 7]. Kesenjangan ini menjadi dasar penelitian ini untuk menguji apakah kualitas komunikatif visual AI dapat diukur secara sistematis.

Penelitian ini bertujuan mengisi kesenjangan tersebut melalui tiga kontribusi utama yang saling melengkapi. Kontribusi pertama adalah pengembangan *Semiotic UX Evaluation Framework* (SUEF), yaitu kerangka evaluasi tujuh dimensi yang mengoperasionalkan konsep-konsep semiotika ke dalam metrik kuantitatif pada protokol *usability testing*. Kerangka ini memungkinkan dimensi semiotik diintegrasikan secara lebih sistematis ke dalam prosedur pengujian UI/UX. Kontribusi kedua adalah pengembangan *pipeline* NLP untuk Bahasa Indonesia melalui integrasi analisis sentimen berbasis IndoBERT dan *topic modeling* menggunakan LDA, yang dikalibrasi untuk kebutuhan evaluasi visual dalam konteks berbahasa Indonesia. Kontribusi ketiga adalah penerapan desain eksperimen komparatif berbasis *eye tracking* dengan membandingkan stimulus visual *AI-generated* dan desain manual dalam sesi *usability testing*. Pendekatan ini memungkinkan perbandingan empiris dan terkontrol terhadap kualitas semiotik dari kedua kategori visual tersebut.

2. TINJAUAN PUSTAKA

Semiotika sebagai kerangka ilmiah untuk memahami makna visual bertumpu pada tiga tradisi utama. Ferdinand de Saussure menjelaskan bahwa tanda terdiri atas penanda (*signifier*), yaitu bentuk material yang dapat dipersepsi, dan petanda (*signified*), yaitu konsep mental yang

diwakilinya. Relasi keduanya bersifat arbitrer dan dibentuk oleh konvensi kultural [1]. Charles Sanders Peirce memperluas konsep ini melalui klasifikasi tanda menjadi ikon, indeks, dan simbol berdasarkan relasinya dengan objek [2]. Roland Barthes menambahkan dimensi kritis melalui perbedaan denotasi sebagai makna literal dan konotasi sebagai lapisan makna yang terbentuk melalui asosiasi kultural maupun ideologis [3]. Dalam Desain Komunikasi Visual, ketiga perspektif ini penting karena keputusan desain, seperti warna, tipografi, simbol, dan komposisi, merupakan keputusan semiotik yang memengaruhi interpretasi pesan oleh audiens.

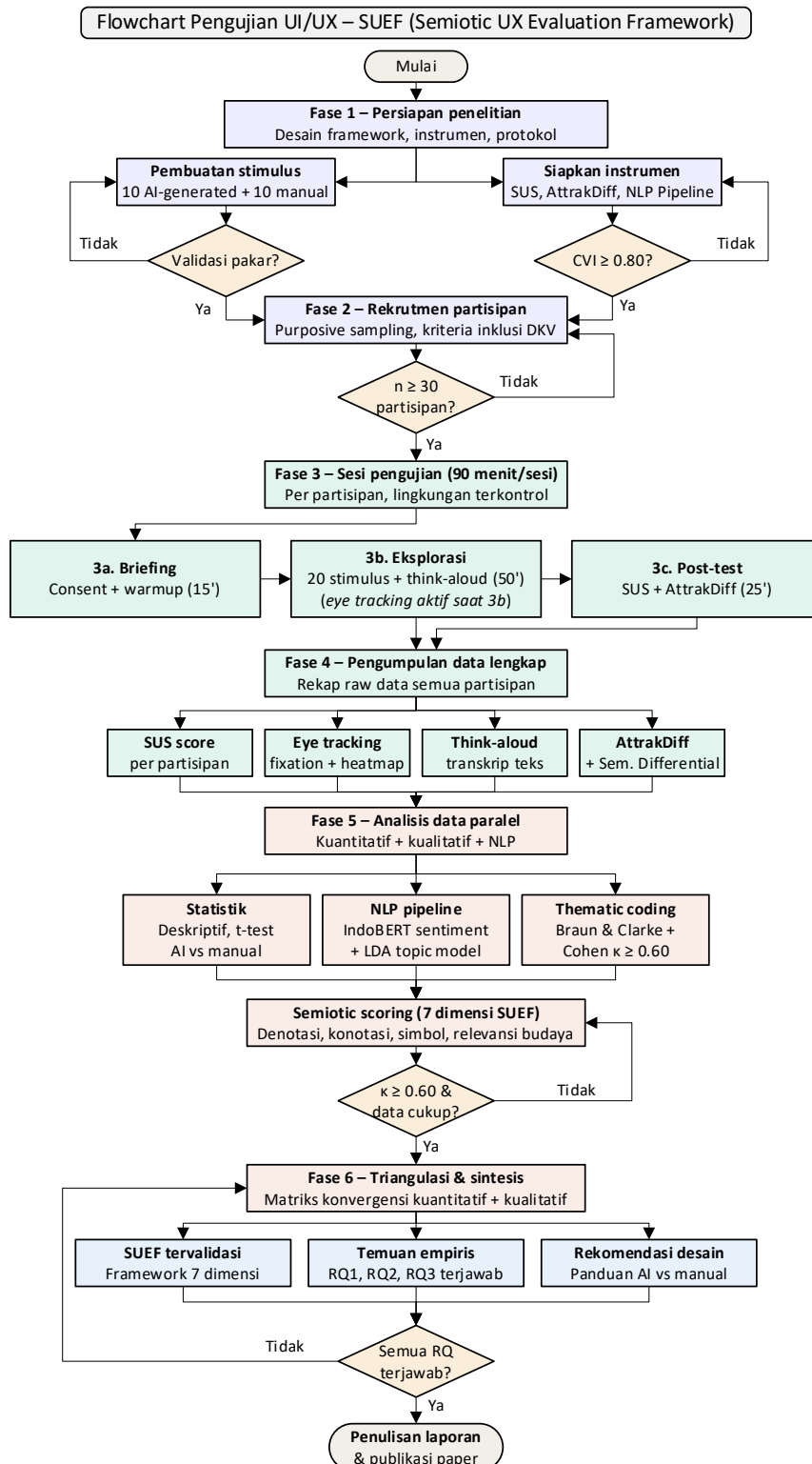
Dalam evaluasi perangkat lunak dan antarmuka pengguna, usability menjadi aspek utama yang perlu diukur secara sistematis. ISO 9241-11 mendefinisikan usability sebagai sejauh mana produk dapat digunakan oleh pengguna tertentu untuk mencapai tujuan tertentu secara efektif, efisien, dan memuaskan dalam konteks penggunaan tertentu [6]. *Usability testing* dilakukan dengan melibatkan pengguna representatif untuk menyelesaikan tugas nyata, sementara perilaku, respons, dan kesulitan mereka diamati serta dianalisis [7]. Salah satu instrumen yang umum digunakan adalah System Usability Scale (SUS), yaitu kuesioner sepuluh item dengan skor 0–100. Skor di atas 68 umumnya dianggap di atas rata-rata, sedangkan skor di atas 80 menunjukkan kategori baik [8, 9]. Pengalaman pengguna yang lebih luas dapat diukur melalui model Hassenzahl yang membedakan kualitas pragmatis dan hedonik, serta dioperasionalkan melalui AttrakDiff [10]. Selain itu, Semantic Differential Scale dapat digunakan untuk mengukur makna konotatif melalui pasangan kata sifat berlawanan pada skala tujuh poin, sehingga menjembatani konsep semiotik abstrak dengan data numerik yang dapat dianalisis secara statistik [11].

Model AI generatif visual seperti Midjourney dan DALL-E umumnya bekerja dengan arsitektur *diffusion model*. Arsitektur ini menghasilkan visual melalui proses iteratif penghilangan *noise* dari gambar acak hingga terbentuk citra yang sesuai dengan instruksi teks pengguna [14]. Dibandingkan pendekatan sebelumnya seperti *generative adversarial network* (GAN), *diffusion model* mampu menghasilkan keluaran yang lebih beragam, stabil, dan mudah dikontrol. Namun, dari perspektif semiotika, kapabilitas estetis ini memiliki keterbatasan mendasar. AI generatif tidak memiliki *intentionality* dalam pengertian filosofis, yaitu kemampuan untuk memiliki intensi dan kesadaran terhadap makna yang hendak dikomunikasikan [4]. Sistem ini bekerja berdasarkan distribusi statistik data pelatihan sehingga visual yang dihasilkan dapat tampak bermakna, tetapi pada dasarnya merupakan penyusunan ulang pola statistik tanpa pemahaman terhadap konteks komunikasi yang sesungguhnya [15].

Keterbatasan tersebut memunculkan fenomena yang dalam penelitian ini disebut *semiotic drift*, yaitu kesenjangan antara makna yang dimaksudkan pengguna melalui *prompt*, makna yang direpresentasikan AI dalam visual, dan makna yang ditafsirkan audiens. Tomalin menunjukkan bahwa sistem AI multimodal dapat menghasilkan deskripsi verbal yang akurat terhadap gambar, tetapi masih kesulitan memahami implikasi semiotik elemen visual dalam konteks sosial nyata [5]. Dalam pendidikan desain, mahasiswa cenderung memosisikan AI sebagai alat bantu eksplorasi visual, bukan sebagai pengganti proses kreatif. Namun, tanpa instrumen evaluasi yang memadai, mereka tetap kesulitan menilai secara sistematis apakah keluaran AI memiliki kualitas komunikatif yang layak [16]. Kondisi ini menunjukkan perlunya kerangka evaluasi yang mengintegrasikan dimensi semiotik ke dalam *usability testing*.

Pemetaan literatur menunjukkan bahwa integrasi semiotika, evaluasi UX, dan AI generatif dalam satu kerangka metodologis masih terbatas. Kajian semiotika visual digital dan studi HCI tentang interaksi dengan alat berbasis AI cenderung berkembang secara terpisah, tanpa jembatan metodologis yang jelas [8, 10]. Sreenivasan dan Suresh menegaskan bahwa integrasi kerangka evaluasi berbasis makna dengan kapabilitas AI generatif merupakan arah riset penting yang perlu dikembangkan [16]. Sejalan dengan itu, Mihalache menunjukkan bahwa pengabaian dimensi semiotik dalam produksi visual AI dapat menyebabkan distorsi makna ketika pesan diterima oleh audiens dari konteks budaya berbeda [22]. Penelitian ini merespons kesenjangan tersebut dengan mengoperasionalkan dimensi semiotik sebagai komponen terukur dalam protokol *usability testing*. Pendekatan ini menawarkan pergeseran metodologis dalam evaluasi HCI dan kajian DKV berbasis AI, karena kualitas komunikatif visual tidak hanya dianalisis secara interpretatif, tetapi juga diukur secara sistematis.

3. METODE PENELITIAN



Gambar 1. Metode penelitian yang digunakan

Penelitian ini menggunakan desain metode campuran (*mixed methods design*) dengan strategi *sequential explanatory*. Pada strategi ini, data kuantitatif dikumpulkan dan dianalisis terlebih dahulu, kemudian diperdalam melalui data kualitatif untuk memberikan interpretasi yang lebih komprehensif

terhadap temuan numerik. Pemilihan desain ini didasarkan pada karakteristik tiga pertanyaan penelitian. RQ1, yang berfokus pada persepsi mahasiswa DKV terhadap usability dan makna semiotik visual *AI-generated* dibandingkan dengan desain manual, memerlukan data perbandingan yang presisi dan dapat diuji secara statistik. RQ2, yang berfokus pada dimensi semiotik yang paling berpengaruh terhadap persepsi makna, memerlukan data perilaku nonverbal yang tidak dapat diperoleh hanya melalui laporan diri responden. RQ3, yang berfokus pada operasionalisasi SUEF sebagai protokol pengujian, memerlukan data untuk memvalidasi kerangka tersebut dari beberapa sudut pandang metodologis secara simultan.

Ketiga kebutuhan tersebut diakomodasi melalui tiga jalur pengumpulan data yang saling melengkapi. Jalur pertama adalah jalur kuantitatif melalui *usability testing* dan kuesioner terstandar, yaitu SUS, AttrakDiff, dan Semantic Differential Scale. Jalur ini menjawab RQ1 dengan menghasilkan skor perbandingan antara visual *AI-generated* dan desain manual yang dapat diuji signifikansinya. Jalur kedua adalah jalur okulomotor melalui perekaman *eye tracking*. Jalur ini menjawab RQ2 dengan mengungkap pola atensi visual partisipan terhadap elemen-elemen semiotik tertentu, sehingga memberikan bukti perilaku yang melengkapi data persepsi dari kuesioner. Jalur ketiga adalah jalur kualitatif melalui *think-aloud protocol* dan wawancara semi-terstruktur. Jalur ini memperdalam pemahaman terhadap RQ1 dan RQ2, sekaligus menyediakan data verbal untuk memvalidasi konsistensi internal dimensi-dimensi SUEF sebagaimana dituntut oleh RQ3. Ketiga jalur tersebut diintegrasikan pada fase triangulasi sebelum disintesis menjadi temuan akhir.

Partisipan penelitian terdiri atas mahasiswa aktif Program Studi DKV semester 3 sampai 6 yang pernah menggunakan setidaknya satu platform AI generatif dalam tiga bulan terakhir. Kriteria ini ditetapkan agar partisipan memiliki pengalaman memadai dalam mengevaluasi visual berbasis AI dan memahami konteks penggunaan AI generatif yang menjadi objek kajian. Ukuran sampel ditetapkan sekurang-kurangnya 30 partisipan dengan mempertimbangkan *statistical power* sebesar 0.80, $\alpha = 0.05$, dan *effect size* sedang (Cohen's $d = 0.5$) pada *paired samples t-test* sebagai analisis utama untuk menjawab RQ1 [7]. Rekrutmen dilakukan melalui *purposive sampling* yang diperluas dengan *snowball sampling*. Seluruh prosedur penelitian telah memperoleh persetujuan komite etika penelitian institusi sebelum pengumpulan data dilaksanakan.

Stimulus pengujian terdiri atas dua set yang dikontrol secara ketat. Set A berisi 10 visual yang dihasilkan menggunakan Midjourney v6, sedangkan Set B berisi 10 visual desain manual yang dibuat oleh desainer berpengalaman. Kedua set mencakup kategori visual yang sama, yaitu identitas merek, poster kampanye, ilustrasi editorial, desain kemasan, dan infografis. Setiap sesi pengujian berlangsung 90 menit secara individual. Partisipan diminta mengeksplorasi 20 stimulus sambil memverbalisasikan proses berpikir melalui *think-aloud protocol*. Setelah itu, partisipan mengisi kuesioner SUS, AttrakDiff, dan *Semantic Differential Scale*. Sesi diakhiri dengan wawancara semi-terstruktur selama 10 menit.

Analisis data dilakukan melalui tiga jalur paralel yang kemudian diintegrasikan melalui triangulasi. Jalur pertama adalah analisis kuantitatif, yang mencakup konversi skor SUS, visualisasi hasil AttrakDiff dalam *portfolio diagram*, serta *paired samples t-test* untuk membandingkan Set A dan Set B disertai pelaporan *effect size*. Jalur kedua adalah analisis NLP terhadap transkripsi *think-aloud* menggunakan *pipeline* berbasis Python. Analisis ini mencakup analisis sentimen dengan IndoBERT dan *topic modeling* menggunakan LDA untuk mengidentifikasi pola makna serta respons verbal partisipan. Jalur ketiga adalah analisis kualitatif menggunakan *thematic analysis* enam fase Braun dan Clarke, yang dilakukan oleh dua peneliti secara independen [12]. Reliabilitas antar-pengkode diukur menggunakan Cohen's Kappa dengan ambang batas $\kappa \geq 0.60$. Temuan yang konsisten pada ketiga jalur analisis dilaporkan sebagai temuan utama, sedangkan temuan yang divergen dianalisis lebih lanjut untuk mengungkap nuansa interpretatif yang tidak dapat dijelaskan oleh satu metode saja.

4. PEMBAHASAN

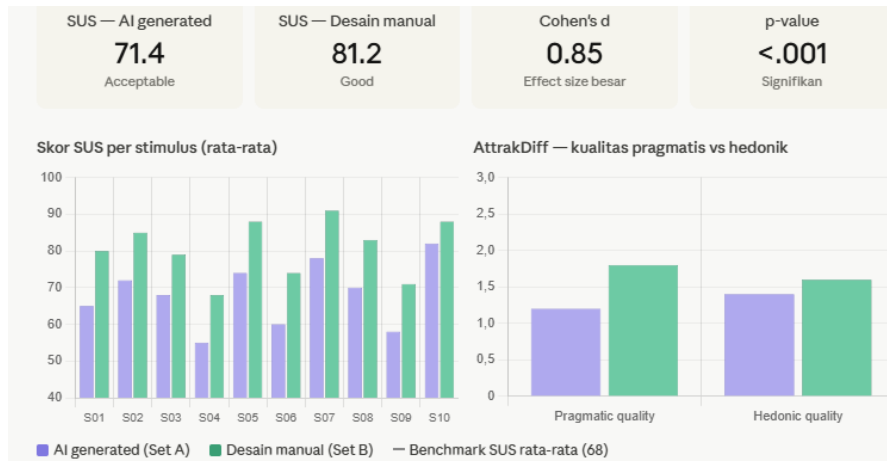
4.1 Profil Partisipan

Penelitian ini melibatkan 32 mahasiswa DKV aktif usia 19–23 tahun. Sebanyak 65.6% yang berada di semester 5 dan 6 dan 34.4% di semester 3 dan 4. Platform AI yang paling banyak digunakan oleh responden pada penelitian ini adalah Midjourney (56.3%), diikuti DALL-E 3 (28.1%) dan Adobe

Firefly (15.6%). Mayoritas partisipan (71.9%) menggunakan platform Generative AI lebih dari sekali seminggu, sehingga dapat diklasifikasikan sebagai pengguna aktif (responden aktif).

4.2 Hasil Usability Testing — SUS dan AttrakDiff

Rata-rata skor SUS Set A (AI-generated) adalah 71.4 (SD = 8.3, kategori acceptable), sedangkan Set B (manual) memperoleh nilai 81.2 (SD = 6.7, kategori good). Perbedaan ini signifikan secara statistik ($t(31) = 4.82$, $p < 0.001$, Cohen's $d = 0.85$). Pada AttrakDiff, Set B unggul terutama pada kualitas pragmatis (1.8 vs 1.2), sementara kualitas hedonik hampir setara (1.6 vs 1.4) hal ini menunjukkan bahwa AI telah mencapai nilai paritas estetis namun masih tertinggal secara komunikasi fungsional.

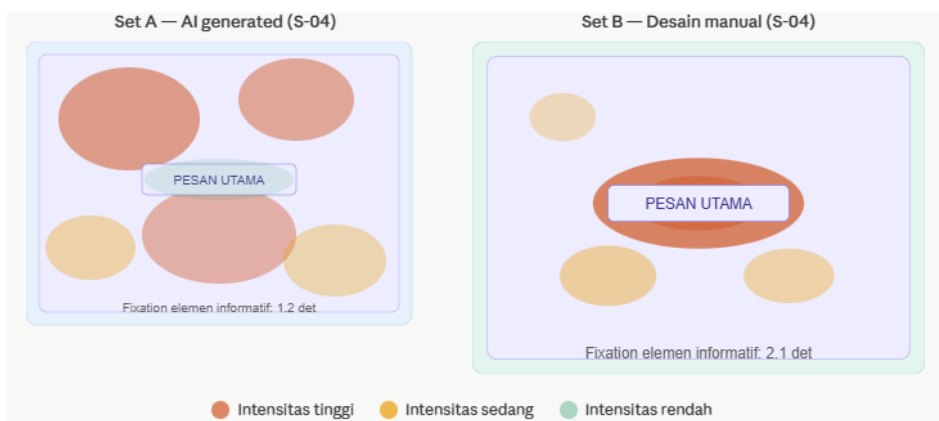


Gambar 2. SUS score

Pada gambar terlihat partisipan P-07 memberikan skor SUS 55 untuk stimulus S-04 (Set A) dan 85 untuk S-04 (Set B). Dalam wawancara P-07 menyatakan visual AI untuk poster kampanye lingkungan "terlihat bagus dan modern, tapi butuh waktu lama untuk paham pesan utamanya." Sebaliknya, poster manual langsung menyampaikan urgensi isu lingkungan melalui hierarki visual yang terstruktur.

4.3 Hasil Eye Tracking sebagai Pola Atensi Visual

Analisis *eye tracking* menunjukkan adanya perbedaan substansial pada *first fixation point* dan *fixation duration* antara Set A dan B. Pada Set A, fiksasi pertama partisipan umumnya tertuju pada elemen dekoratif (gradasi warna, tekstur, dan ornamen) sebelum berpindah ke elemen informatif. Rata-rata *fixation duration* pada elemen yang memuat pesan utama sebesar 0.4 detik untuk Set A dan 2.8 detik untuk Set B. Temuan ini menunjukkan bahwa visual *AI-generated* cenderung lebih cepat menarik perhatian pada aspek dekoratif, sedangkan desain manual lebih efektif mengarahkan perhatian partisipan pada elemen pesan utama.

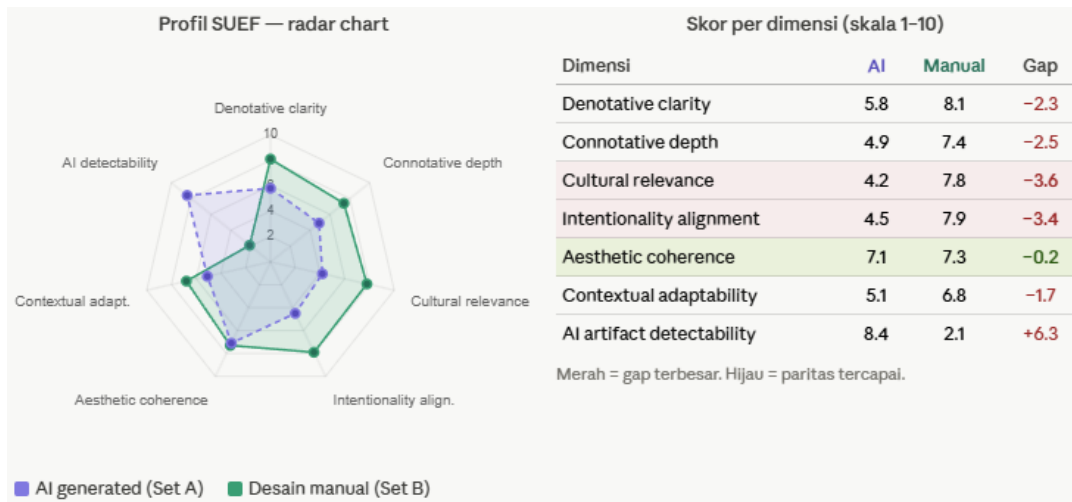


Gambar 3 Data eye tracking

Pola heatmap di atas merepresentasikan temuan pada stimulus S-04 (infografis konsumsi plastik). Pada Set A, fiksasi intens tersebar di area latar belakang dekoratif yang estetis, sementara angka-angka data utama hanya mendapat rata-rata 0.4 detik. Pada Set B, fiksasi langsung terkonsentrasi pada elemen informatif utama (rata-rata 2.8 detik) sebelum mengikuti alur infografis yang terstruktur.

4.4 Semiotic Scoring — Tujuh Dimensi SUEF

Penilaian tujuh dimensi SUEF melalui Semantic Differential Scale dan pengkodean tematik menghasilkan profil yang jelas. Dua dimensi dengan selisih terbesar adalah *Cultural Relevance* (−3.6) dan *Intentionality Alignment* (−3.4), keduanya berkaitan langsung dengan keterbatasan fundamental AI. Sebaliknya, *Aesthetic Coherence* hampir tidak berbeda (−0.2), mengkonfirmasi paritas estetis AI.

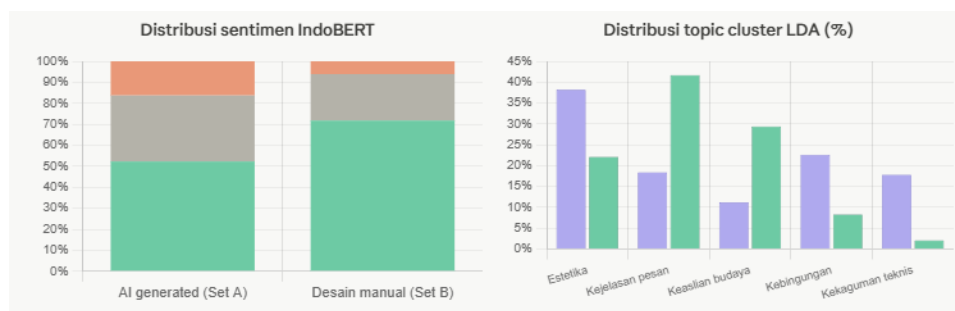


Gambar 4 Nilai *semiotic scoring*

Contoh konkret terlihat pada stimulus S-09 berupa poster bertema “Keberagaman Budaya Indonesia”. Visual yang dihasilkan Midjourney menampilkan campuran elemen Asia Tenggara yang bersifat generik, termasuk kombinasi motif yang menyerupai budaya Thailand dan Malaysia. Kondisi ini menyebabkan skor *Cultural Relevance* pada Set A hanya mencapai 4.2. Partisipan P-19 dalam sesi *think-aloud* menyatakan, “Ini kesannya budaya Asia Tenggara secara umum, tidak ada yang benar-benar khas Indonesia.” Sebaliknya, desain manual menggunakan elemen budaya yang lebih spesifik, seperti tenun ikat Sumba, topeng Malang, dan pola kawung Jawa. Elemen-elemen tersebut lebih mudah dikenali oleh partisipan dan menghasilkan skor *Cultural Relevance* sebesar 8.3.

4.5 Hasil NLP Pipeline

Analisis sentimen IndoBERT terhadap 2.847 segmen transkripsi mengungkapkan *distribusi* yang berbeda signifikan antara kedua set. Tema yang mendominasi respons terhadap Set A adalah estetika dan kekaguman teknis, sementara Set B didominasi tema kejelasan pesan dan relevansi kontekstual.



Gambar 5. NLP pipeline

Sentimen positif sebesar 19.5 poin persentase antara kedua set signifikan secara statistik ($\chi^2 = 47.3, p < 0.001$). Topic cluster "kekaguman teknis" hampir eksklusif muncul pada Set A (17.8% berbanding 2.1%), mencerminkan bahwa partisipan mengagumi kemampuan rendering AI tetapi tidak kualitas komunikatifnya.

4.6 Pembahasan Integratif

Sintesis dari keempat jalur data *usability testing*, *eye tracking*, *semiotic scoring*, dan NLP *pipeline* menghasilkan tiga temuan integratif yang saling mengkonfirmasi dan secara bersama-sama menjawab pertanyaan penelitian yang telah dirumuskan. Berkaitan dengan permasalahan pertama, temuan penelitian ini menunjukkan bahwa mahasiswa DKV secara konsisten mempersepsikan *visual AI-generated* sebagai produk yang unggul secara estetis namun kurang efektif secara komunikatif.

Rata-rata skor SUS sebesar 71.4 menempatkan visual AI pada kategori *acceptable*. Skor ini menunjukkan bahwa pengguna masih dapat berinteraksi dengan visual tersebut tanpa hambatan teknis yang berarti, tetapi kualitas usabilitynya belum mencapai kategori *good* sebagaimana ditunjukkan oleh desain manual dengan skor 81.2. Data *eye tracking* memperjelas mekanisme yang mendasari perbedaan tersebut. Partisipan yang mengevaluasi Set A menunjukkan pola fiksasi yang lebih tersebar dan waktu pencarian elemen informatif yang lebih panjang. Pola ini mengindikasikan adanya beban kognitif yang lebih tinggi dalam proses ekstraksi makna. Temuan tersebut menunjukkan bahwa visual *AI-generated* cenderung menuntut upaya interpretatif yang lebih besar dari pengguna untuk memahami pesan yang seharusnya dapat disampaikan secara efisien melalui hierarki visual yang terstruktur.

Berkaitan dengan permasalahan kedua, analisis terhadap tujuh dimensi SUEF menunjukkan bahwa *Cultural Relevance* dan *Intentionality Alignment* merupakan dimensi yang paling menentukan perbedaan kualitas UX antara kedua set stimulus, dengan selisih masing-masing sebesar 3.6 dan 3.4 poin. Kedua dimensi tersebut memiliki karakteristik yang sama, yaitu bergantung pada pemahaman terhadap konteks sosial, tujuan komunikasi, serta nuansa budaya yang spesifik. Kemampuan ini tidak dimiliki secara fundamental oleh sistem AI generatif yang bekerja berdasarkan pencocokan pola statistik. Temuan ini memperkuat argumen teoretis dalam tinjauan pustaka mengenai ketiadaan *intentionality* pada AI [4, 15]. Sebaliknya, dimensi *Aesthetic Coherence* hanya menunjukkan selisih 0.2 poin antara kedua set stimulus. Hasil ini mengindikasikan bahwa model AI generatif telah mendekati paritas estetis dengan desain manual. Implikasinya, evaluasi kualitas visual berbasis AI tidak lagi memadai jika hanya mengandalkan penilaian estetis, tetapi perlu dilengkapi dengan instrumen yang secara khusus mengukur dimensi komunikatif dan semiotik.

Berkaitan dengan permasalahan ketiga, penelitian ini menunjukkan bahwa SUEF dapat dioperasionalkan sebagai protokol pengujian berbasis perangkat lunak. Reliabilitas antar-pengkode pada pengkodean tematik mencapai $\kappa = 0.71$, melampaui ambang batas yang ditetapkan ($\kappa \geq 0.60$), sehingga mengindikasikan konsistensi yang memadai dalam penerapan dimensi-dimensi *framework* oleh peneliti yang berbeda. Selain itu, profil *semiotic scoring* yang dihasilkan menunjukkan pola yang konsisten lintas empat metode yang digunakan. Temuan dari *Semantic Differential Scale*, data *eye tracking*, analisis sentimen berbasis NLP, dan *thematic analysis* secara umum mengarah pada kesimpulan yang sama, yaitu keunggulan stimulus AI pada dimensi estetis serta kelemahannya pada dimensi kultural-komunikatif. Konsistensi lintas metode ini memberikan dukungan terhadap validitas konvergen SUEF dan menunjukkan bahwa strategi triangulasi dalam desain penelitian mampu menghasilkan pembacaan yang saling menguatkan.

Secara keseluruhan, temuan penelitian ini menegaskan bahwa perbedaan kualitas antara visual *AI-generated* dan desain manual tidak terutama terletak pada aspek teknis maupun estetis, melainkan pada dimensi semiotik yang lebih fundamental. Perbedaan tersebut tampak pada kemampuan visual dalam membangun makna yang relevan secara kultural, spesifik secara kontekstual, dan selaras dengan intensi komunikatif yang ingin disampaikan. Dimensi ini sulit dijangkau oleh instrumen evaluasi UX konvensional apabila evaluasi hanya berfokus pada

efisiensi, kemudahan penggunaan, dan kepuasan pengguna. Dengan demikian, temuan ini tidak hanya memperkuat kebutuhan akan framework seperti SUEF, tetapi juga membuka arah penelitian baru yang mengintegrasikan teori semiotika ke dalam praktik evaluasi HCI, khususnya sebagai respons terhadap meluasnya adopsi AI generatif dalam industri desain komunikasi visual.

5. KESIMPULAN

Penelitian ini mengembangkan dan memvalidasi *Semiotic UX Evaluation Framework* (SUEF) sebagai respons terhadap gap dalam literatur HCI, yaitu belum tersedianya instrumen evaluasi yang secara eksplisit mengintegrasikan dimensi semiotika ke dalam protokol *usability testing* untuk menilai kualitas komunikatif visual berbasis AI. Melalui desain metode campuran yang melibatkan 32 mahasiswa DKV sebagai partisipan, penelitian ini menguji SUEF secara empiris dengan menggabungkan *usability testing* (SUS dan AttrakDiff), *eye tracking*, *Semantic Differential Scale*, serta pipeline NLP berbasis IndoBERT dan LDA. Hasil penelitian menunjukkan bahwa SUEF dapat dioperasionalkan sebagai protokol pengujian yang valid dan reliabel. Hal ini ditunjukkan oleh reliabilitas antar-pengkode yang mencapai $\kappa = 0.71$ serta profil *semiotic scoring* yang menunjukkan konsistensi lintas keempat jalur analisis.

Temuan empiris penelitian ini menunjukkan bahwa visual *AI-generated* memperoleh skor usability dalam kategori *acceptable* (SUS = 71.4), tetapi belum mencapai kategori *good* sebagaimana ditunjukkan oleh desain manual (SUS = 81.2). Perbedaan tersebut signifikan secara statistik dengan *effect size* besar (Cohen's $d = 0.85$). Data *eye tracking* mengungkap mekanisme yang mendasari perbedaan ini. Pengguna menunjukkan beban kognitif yang lebih besar ketika mengekstraksi makna dari visual AI karena hierarki visualnya kurang terstruktur secara semiotik. Hal ini terlihat dari pola fiksasi yang lebih tersebar dan durasi pencarian elemen informatif yang lebih panjang. Analisis NLP terhadap lebih dari 2.800 segmen transkripsi *think-aloud* memperkuat pola tersebut. Respons terhadap Set A didominasi oleh tema estetika dan kekaguman teknis, sedangkan respons terhadap Set B lebih banyak berkaitan dengan kejelasan pesan dan relevansi kontekstual.

Temuan terpenting penelitian ini adalah bahwa ketimpangan antara visual *AI-generated* dan desain manual tidak terutama bersifat estetis. Dimensi *Aesthetic Coherence* hanya menunjukkan selisih 0.2 poin, sehingga mengindikasikan bahwa keduanya telah mendekati paritas estetis. Kesenjangan utama justru bersifat semiotik, terutama pada dimensi *Cultural Relevance* dengan selisih 3.6 poin dan *Intentionality Alignment* dengan selisih 3.4 poin. Kedua dimensi tersebut bergantung pada pemahaman terhadap konteks budaya dan intensi komunikatif yang spesifik, yaitu kemampuan yang tidak dimiliki secara fundamental oleh sistem AI yang bekerja berdasarkan distribusi statistik data pelatihan. Temuan ini menunjukkan bahwa AI generatif telah mampu menghasilkan visual yang tampak baik secara estetis, tetapi belum mampu memastikan bahwa visual tersebut mengomunikasikan makna secara tepat dalam konteks sosial dan budaya tertentu.

Penelitian ini memberikan kontribusi pada tiga aspek utama. Secara teoretis, penelitian ini menunjukkan bahwa konsep-konsep semiotika dapat dioperasionalkan sebagai dimensi pengukuran kuantitatif yang valid dalam evaluasi HCI. Secara metodologis, *pipeline mixed methods* yang dikembangkan menghasilkan proses triangulasi yang koheren dan dapat direplikasi pada konteks penelitian lain. Secara praktis, SUEF menyediakan panduan berbasis bukti bagi tim desain dan pendidik DKV untuk menentukan kapan intervensi manusia diperlukan dalam memastikan kualitas komunikatif keluaran AI. Meskipun sampel penelitian ini masih terbatas pada satu institusi dan replikasi lintas konteks tetap diperlukan, hasil penelitian ini menegaskan bahwa evaluasi visual berbasis AI tidak cukup hanya menilai kualitas estetis. Pada era ketika mesin dapat menghasilkan visual yang menarik dalam waktu singkat, pertanyaan utama bukan lagi apakah AI dapat menghasilkan gambar yang baik, melainkan apakah gambar tersebut benar-benar mampu berkomunikasi. SUEF menjadi salah satu langkah awal untuk menjawab pertanyaan tersebut secara lebih terukur.

DAFTAR PUSTAKA

- [1] F. De Saussure, *Course in General Linguistics*. New York: McGraw-Hill Book Company, 1996.
- [2] R. M. Gazoni, "A Semiotic Analysis of Programming Languages," pp. 91–101, 2018, doi: 10.4236/jcc.2018.63007.
- [3] Y. Qian, Q. Bao, S. Zhang, and X. Peng, "A cultural memory semiotics and function behavior structure model for digital inheritance and innovation in AI generated Huizhou woodcarving images," *Sci. Rep.*, 2026, doi: <https://doi.org/10.1038/s41598-026-35360-5>.
- [4] M. Tomalin, "Multimodal social semiotics and the challenge of artificial intelligence," *Multimodality Soc.*, vol. 5, no. 2, pp. 223–244, 2025, doi: 10.1177/26349795251327939.
- [5] M. G. Dondero, *Semiotics of artificial intelligence : enunciative praxis in image analysis and generation*. 2025.
- [6] M. Cang, *Semiotics-AHP-Generative AI Pipeline for Intelligent Generation and Design of Shima Stone-Carving Motifs*, vol. 1, no. 1. Association for Computing Machinery, 2023.
- [7] R. Berthes, *Mythologies*. New York, 1972.
- [8] M. Thellefsen, B. Sørensen, and A. N. Dewi, "Generative AI and the semiotic reconfiguration of knowledge organization – a preliminary exploration," *J. Doc.*, vol. 81, pp. 1181–1199, Aug. 2025, doi: 10.1108/JD-04-2025-0094.
- [9] L. Macnaught, "Customising chatbots for writing development: Anticipating semiotic mediation with the theoretical architecture of systemic functional linguistics Lucy," *J. English Acad. Purp.*, vol. 80, no. October 2025, p. 101646, 2026, doi: 10.1016/j.jeap.2026.101646.
- [10] R. V. Kozinets, "Netnography: Understanding Networked Communication Society," in *Handbook of Social Media Research Method*, Toronto: York University, 2022, pp. 1–26.
- [11] J. W. Creswell, *Research Design, Qualitative, Quantitative, and Mixed Method Approaches*. California: SAGE Publications Inc., 2014.
- [12] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006.
- [13] S. J. Russell *et al.*, *Artificial Intelligence: A Modern Approach*. New Jersey: Alan Apt, 2021.
- [14] Y. K. Dwivedi *et al.*, "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, no. March, 2023, doi: 10.1016/j.ijinfomgt.2023.102642.
- [15] A. Sreenivasan and M. Suresh, "International Journal of Innovation Studies Design thinking and artificial intelligence : A systematic literature review exploring synergies," *Int. J. Innov. Stud.*, vol. 8, no. 3, pp. 297–312, 2024, doi: 10.1016/j.ijis.2024.05.001.
- [16] L. Floridi, *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*, vol. 6649, no. 2023. 2025.
- [17] R. Mihalache, "Artificial Intelligence Meets Semiotics : Optimising Advertising Strategies across Cultures," in *Sciendo*, 2025, pp. 4652–4670, doi: 10.2478/picbe-2025-0355.

Biodata Penulis

Nashiruddin Alfath, S.Sn., M.Sn., Lahir di Jakarta. Meraih gelar Sarjana Seni di Program Studi Animasi pada Fakultas Disain Komunikasi Visual tahun 2016. Meraih Magister Seni (MSn), Penciptaan Seni, di Institut Kesenian Jakarta tahun 2018. Saat ini penulis berprofesi sebagai dosen LLDIKTI Wilayah III dipekerjakan pada Universitas Bina Nusantara, Fakultas Desain Komunikasi Visual.

Prof Dr. Yufiarti, M.Psi. Lahir di Jakarta. Meraih sarjana Psikologi Pendidikan di IKIP Jakarta (Sekarang Universitas Negeri Jakarta) tahun 1986. Magister Psikologi di Universitas Indonesia Fakultas Psikologi tahun 1991. Doktor Pendidikan di IKIP Jakarta (UNJ) tahun 1996. Saat ini berprofesi sebagai dosen di Fakultas Psikologi, Universitas Negeri Jakarta.