

Implementasi konsep *tabayyun* dan fatwa MUI dalam perancangan sistem verifikasi fakta era kecerdasan buatan

Hastuti^{1*}, Imam Mubarak A¹, Ryan Andika¹

¹Program Studi Sistem Informasi

Fakultas Sains dan Teknologi, Universitas Islam Negeri Alauddin Makassar

Jl. H. M. Yasin Limpo No. 36, Gowa, Sulawesi Selatan, Indonesia. 92118

*E-mail: hastuti.baharuddin@uin-alauddin.ac.id

Abstrak: Perkembangan teknologi Kecerdasan Buatan (AI) generatif telah mengakselerasi penyebaran disinformasi seperti *deepfake* secara masif, sehingga memicu krisis etika di ruang digital. Penelitian ini bertujuan untuk menjembatani kesenjangan sosio-teknikal antara regulasi etis keagamaan dan rekayasa perangkat lunak dalam menangani masalah tersebut. Metodologi yang diterapkan adalah kualitatif deskriptif-analitis dengan pendekatan studi kepustakaan (*library research*). Kajian terdahulu umumnya terisolasi pada dua kutub yang terpisah, yakni kajian teologis-normatif murni yang hanya merekomendasikan literasi kognitif, atau rekayasa algoritma pendeteksi hoaks tanpa menyertakan variabel kepatuhan etika pengguna. Berbeda dengan penelitian sebelumnya, diargumenkan bahwa doktrin etika Islam tidak boleh hanya menjadi imbauan moral yang pasif. Oleh karena itu, kebaruan (*novelty*) dari penelitian ini terletak pada gagasan orisinal yang memposisikan teks Fatwa MUI No. 24 Tahun 2017 dan Fiqih Informasi sebagai *System Requirements Specification* (SRS) dan logika bisnis (*business rules*) dalam arsitektur perangkat lunak. Melalui sintesis antardisiplin ini, dihasilkan sebuah model arsitektur konseptual murni tanpa pengujian empiris langsung, yang menyimpulkan bahwa penggabungan rekayasa antarmuka berupa *interface friction* dan otomatisasi pemrosesan bahasa model BERT berbasis nilai *tabayyun* terbukti secara konseptual dapat mengekang impulsivitas pengguna dan mereduksi penyebaran misinformasi generatif secara preventif.

Kata Kunci: Fatwa MUI, fiqih informasi, kecerdasan buatan, sistem verifikasi fakta, *tabayyun*

Abstract: Implementation of Tabayyun Concept and MUI Fatwa in Designing Fact-Checking Systems in the Artificial Intelligence Era. The development of generative Artificial Intelligence (AI) technology has massively accelerated the spread of disinformation such as deepfakes, triggering an ethical crisis in the digital space. This study aims to bridge the socio-technical gap between religious ethical regulations and software engineering in addressing this issue. The methodology applied is descriptive-analytical qualitative research with a library research approach. Previous studies are generally isolated into two separate poles: purely theological-normative studies that only recommend cognitive literacy, or hoax detection algorithm engineering without including user ethical compliance variables. Unlike previous studies, it is argued that Islamic ethical doctrines should not remain passive moral appeals. Therefore, the novelty of this research lies in the original idea of positioning the text of MUI Fatwa No. 24 of 2017 and Information Fiqh as System Requirements Specifications (SRS) and business rules within software architecture. Through this interdisciplinary synthesis, a purely conceptual architecture model is produced without direct empirical testing, which concludes that combining interface friction engineering and BERT language processing automation based on *tabayyun* values conceptually restrains user impulsivity and reduces the spread of generative misinformation more preventively.

Keywords: MUI fatwa, information fiqh, artificial intelligence, fact-checking system, *tabayyun*

PENDAHULUAN

Perkembangan pesat teknologi kecerdasan buatan (*Artificial Intelligence/AI*) generatif telah memicu pergeseran paradigma dalam ekosistem komunikasi digital global. Akselerasi ini memicu fenomena *information disorder* yang terotomatisasi secara masif. Kehadiran teknologi manipulasi digital terkini seperti *Generative Adversarial Networks* (GANs) yang memproduksi *deepfake* telah menciptakan disinformasi yang sangat realistis (Nurdin & Nugraha, 2025). Terdapat sebuah paradoks yang berbahaya di sini: disinformasi diproduksi secara eksponensial oleh mesin super cepat, namun mekanisme perlawanan masih mengandalkan cara yang sangat lambat, yakni pengecekan fakta secara manual oleh manusia. Dampak dari asimetri ini sangat destruktif, mulai dari polarisasi sosial hingga erosi kepercayaan publik terhadap institusi resmi di Indonesia (Mastori et al., 2025).

Kerentanan infrastruktur digital ini diperparah oleh kondisi sosiologis pengguna media sosial di Indonesia yang sedang mengalami krisis etika komunikasi. Ruang digital sering kali dieksploitasi untuk menyebarkan konten provokatif yang secara nyata menegasi nilai-nilai akhlak Islam (Zaini et al., 2026). Fenomena ini tidak hanya menjangkiti masyarakat awam, tetapi juga kalangan terpelajar yang kerap kehilangan nalar kritis akibat paparan algoritma media sosial yang memang dirancang untuk memanipulasi emosi. Akibatnya, netizen dengan sangat mudah diorkestrasi menjadi agen penyebar disinformasi secara massal.

Sebagai respons strategis terhadap kegentingan ini, Majelis Ulama Indonesia mengeluarkan Fatwa MUI No. 24 Tahun 2017 tentang Hukum dan Pedoman Bermuamalah melalui Media Sosial (Lembaga Fatwa Majelis Ulama Indonesia, 2024). Secara komprehensif, fatwa ini mengharamkan aktivitas memproduksi, menyebarkan, dan memfasilitasi hoaks, fitnah, ghibah, namimah, hingga aktivitas buzzer komersial (Ngarifin & Sutopo, 2025). Fatwa ini secara teologis mengaktualisasikan perintah *tabayyun* dalam Surah Al-Hujurat ayat 6, yang mewajibkan pengujian terhadap validitas sumber berita (*sanad*) dan kebenaran konten (*matan*). Pandangan ini sejalan dengan konsolidasi fikih informasi kontemporer yang memposisikan hoaks sebagai pelanggaran sistemik terhadap kemaslahatan umat (Hastuti et al., 2023).

Meskipun instrumen normatif ini sangat kuat, secara kritis implementasinya di lapangan dinilai menemui jalan buntu. Hukum positif seperti UU ITE hanya berfungsi pasca-kejadian (menghukum setelah hoaks viral), sementara imbauan moral keagamaan sering kali diabaikan di ruang siber. Kajian akademik terdahulu pun cenderung terjebak pada dikotomi yang kaku. Para sarjana agama terus menggaungkan *tabayyun* hanya sebagai literasi kognitif tanpa menawarkan solusi sistemik, sementara ahli ilmu komputer (IT) sibuk merancang algoritma deteksi hoaks seperti *Natural Language Processing* (NLP) BERT tanpa membekalinya dengan kerangka etika sosial-keagamaan yang sesuai dengan kultur masyarakat penggunanya (Fardhina et al., 2025).

Berangkat dari kelemahan riset-riset tersebut, gagasan utama yang ditawarkan dalam artikel ini adalah sebuah integrasi radikal yang disebut *interdisciplinary software engineering*. Diargumenkan bahwa Fatwa MUI No. 24 Tahun 2017 tidak boleh lagi diperlakukan sekadar sebagai teks ceramah keagamaan. Fatwa ini harus "dihidupkan" dengan cara mengubah poin-poin hukumnya menjadi Spesifikasi Kebutuhan Sistem (*System Requirements Specification / SRS*) dalam merancang Sistem Verifikasi Fakta.

Oleh karena itu, rumusan masalah dalam penelitian ini difokuskan pada parameter etis *tabayyun* (Fatwa MUI No. 24 Tahun 2017 dan Fiqih Informasi) ditransformasikan

menjadi *System Requirements* yang dapat dipahami oleh mesin serta model arsitektur *back-end* dan rancangan antarmuka (UI/UX) dibangun agar mampu "memaksa" penegakan fikih informasi tersebut secara otomatis.

Hipotesis dan klaim baru yang diajukan adalah penanggulangan hoaks generatif AI tidak akan pernah berhasil jika hanya mengandalkan kecerdasan algoritma pendeteksi di belakang layar (*back-end*). Pemberantasan hoaks harus dilakukan dengan merekayasa perilaku pengguna secara langsung pada layar sentuh mereka (*front-end*) melalui penerapan hambatan antarmuka (*interface friction*). Konsep hambatan antarmuka ini didukung oleh kajian *Human-Computer Interaction* (HCI) kontemporer sebagai bentuk perlawanan terhadap dark patterns yang efektif untuk memicu jeda kognitif pengguna (Cox et al., 2016; Gray et al., 2018). Dengan menanamkan nilai tabayyun langsung ke dalam kode pemrograman (*ethical-by-design*), sistem dapat meredam impulsivitas pengguna secara preventif.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif deskriptif-analitis dengan metode studi kepustakaan (*library research*). Pemilihan metode ini didasarkan pada kebutuhan penulis untuk menyintesis dua domain ilmu yang berbeda: teks normatif keagamaan dan teori rekayasa perangkat lunak (*software engineering*). Objek formal penelitian ini adalah konsep perancangan arsitektur sistem informasi, sedangkan objek materialnya adalah teks Fatwa MUI No. 24 Tahun 2017 serta literatur fikih informasi.

Agar tidak hanya berhenti pada analisis normatif, metode penelitian ini diperluas menggunakan kerangka kerja Perancangan Sistem Konseptual (*Conceptual System Design*). Rangkaian perancangan sistem informasi dalam penelitian ini dilakukan melalui tiga tahapan operasional yang terstruktur:

1. Tahap *Requirement Elicitation* (Elisitasi Kebutuhan): Menggunakan metode analisis isi (*content analysis*) yang diintegrasikan dengan pemodelan konseptual sistem, dalil-dalil hukum dari Fatwa MUI No. 24 Tahun 2017 diekstraksi dan direduksi menjadi variabel terukur. Instrumen utama yang digunakan dalam proses elisitasi ini adalah matriks kodifikasi etis (lihat Tabel 1). Melalui instrumen ini, parameter larangan (seperti ghibah, namimah) dipetakan menjadi batasan fungsional (*negative use cases*), sedangkan parameter anjuran (seperti *tabayyun*) ditransformasikan menjadi parameter validasi wajib (*mandatory validation checks*). Hasil ekstraksi ini kemudian diklasifikasikan ke dalam Spesifikasi Kebutuhan Sistem (*System Requirements Specification / SRS*).

Tabel 1. Matriks pemetaan teks fatwa MUI ke spesifikasi kebutuhan sistem

Teks Fatwa MUI / Fikih Informasi	Teks Fatwa MUI / Fikih Informasi	Teks Fatwa MUI / Fikih Informasi	Teks Fatwa MUI / Fikih Informasi
Larangan menyebarkan dan memfasilitasi hoaks/fitnah	<i>Ghibah / Namimah</i>	Sistem wajib mencegah (<i>intercept</i>) transmisi teks yang terindikasi sentimen adu domba.	<i>Negative Use Case / Functional Constraint</i>
Kewajiban menguji validitas dan kebenaran sumber	<i>Tabayyun (Sanad)</i>	Sistem secara otomatis melacak reputasi histori URL/Domain pada API <i>Database Nasional</i> .	<i>Mandatory Validation Check</i>
Larangan melakukan dekontekstualisasi fakta masa lalu	Fitnah Kubra	Sistem mengekstraksi dan mencocokkan <i>timestamp</i> metadata dengan konteks teks.	<i>Cross-Reference Verification</i>

2. Tahap *System Workflow Modeling* (Pemodelan Alur Sistem): Merancang arsitektur aliran data (*data flow*) secara berurutan, mulai dari saat informasi didistribusikan oleh pengguna, dianalisis keabsahannya oleh kecerdasan buatan (algoritma NLP BERT) di server (*back-end*), hingga sistem menghasilkan keputusan verifikasi secara komputasional (*use case workflow*).
3. Tahap *UI/UX Intervention Design* (Perancangan Intervensi Antarmuka): Menerjemahkan sanksi moral keagamaan dan etika *Qaulan Baligha* ke dalam purwarupa rekayasa hambatan visual (*interface friction*) yang akan diaplikasikan dan membatasi impulsivitas langsung pada layar perangkat pengguna (*front-end*).
4. Tahap Rencana Evaluasi Sistem: Menetapkan kerangka validasi di masa depan sebagai tindak lanjut dari perancangan konseptual ini. Rencana evaluasi difokuskan pada pengujian tingkat kebergunaan antarmuka menggunakan *System Usability Scale* (SUS) dan validasi performa algoritma pendeteksi hoaks model NLP BERT menggunakan metrik pengukuran *F1-Score*.

HASIL DAN PEMBAHASAN

Rekonstruksi Sistem Verifikasi Fakta sebagai *Business Rules*

Berdasarkan pendekatan *Conceptual System Design*, penelitian ini menghasilkan sebuah rumusan *System Requirements Specification* (SRS) yang secara presisi mentransformasikan hukum normatif agama menjadi arsitektur komputasional terukur. Untuk memastikan keandalan sistem, rancangan spesifikasi ini disusun agar memenuhi parameter *Quality Assurance* (QA) perangkat lunak yang ketat dengan arsitektur berorientasi layanan (*Microservices Architecture*). Penulis memetakan spesifikasi fungsional *back-end* ke dalam tiga modul inti yang beroperasi secara paralel:

1. Modul Verifikasi *Sanad* (*Digital Provenance & Node Reputation*): Modul ini tidak sekadar membaca nama akun, melainkan membedah rekam jejak (*log history*) dan struktur *Uniform Resource Locator* (URL). Sistem menjalankan *skrip crawling* proaktif untuk mengalkulasi skor reputasi entitas pengirim melalui integrasi *Application Programming Interface* (API) tersandi (menggunakan enkripsi AES-256) yang terhubung dengan *whitelist* dan *blacklist* basis data Satgas Anti-Hoaks Nasional. Sistem mengevaluasi metrik sentralitas jaringan: apabila sebuah konten didistribusikan secara masif dalam waktu kurang dari 5 menit oleh simpul-simpul akun bot nir-identitas (anonim) yang riwayatnya dipenuhi anomali, sistem secara logis melabeli entitas tersebut sebagai *sanad dhaif* (lemah) atau *maudhu'* (palsu).
2. Modul Verifikasi Matan (Integritas Konten via NLP): Memanfaatkan kecerdasan buatan berbasis *Natural Language Processing* menggunakan arsitektur *Bidirectional Encoder Representations from Transformers* (BERT). Berbeda dengan sistem *rule-based* primitif yang hanya mencocokkan kata kunci kasar, BERT di-*fine-tuning* menggunakan korpus data multibahasa (termasuk dialek lokal, bahasa slang, dan ujaran kebencian berbalut agama). Melalui proses tokenisasi dan ekstraksi *contextual embeddings* dua arah (membaca kalimat dari kiri ke kanan dan sebaliknya), sistem memiliki kepekaan sintaksis untuk membedakan sarkasme murni dengan probabilitas namimah (adu domba) berisiko tinggi. Hal ini menekan tingkat *false positive* secara signifikan. Namun, tantangan teknis krusial dalam tahap *fine-tuning* model BERT ini adalah tingginya variasi bahasa gaul (*slang*), penyingkatan kata ekstrem, dan struktur kalimat informal di Indonesia yang tergolong *low-resource language*, sehingga implementasinya memerlukan anotasi korpus data linguistik lokal yang ekstensif secara berkala.

3. Modul Verifikasi Konteks (Audit Telemetry & Forensik Digital): Dirancang untuk menangkal disinformasi visual. Karena manipulator sering menghapus (*stripping*) metadata asli, modul ini tidak hanya bergantung pada *Exchangeable Image File Format* (EXIF) data (seperti *timestamp* dan *geotagging* spasial), tetapi juga mengimplementasikan algoritma *Perceptual Hashing* (pHash). Sistem menghasilkan "sidik jari digital" dari suatu gambar/video dan membandingkannya dengan *database* hoaks lama. Sistem kemudian memvalidasi silang *timestamp* forensik tersebut dengan klaim narasi temporal yang ditulis pengguna, guna mengungkap taktik dekontekstualisasi (video masa lalu dinarasikan sebagai tragedi hari ini).

Visualisasi Arsitektur dan Rangkaian Alur Sistem

Rancangan *Tabayyun* Sistemik ini divisualisasikan ke dalam sebuah rangkaian kerja terdistribusi yang dirancang dengan target *Non-Functional Requirement* (NFR) berupa latensi di bawah 2 detik per *request*. Alur sistem terbagi menjadi tiga lapisan (*layer*) utama:

1. Layer 1: Input & Data Acquisition (*Client-Side*)

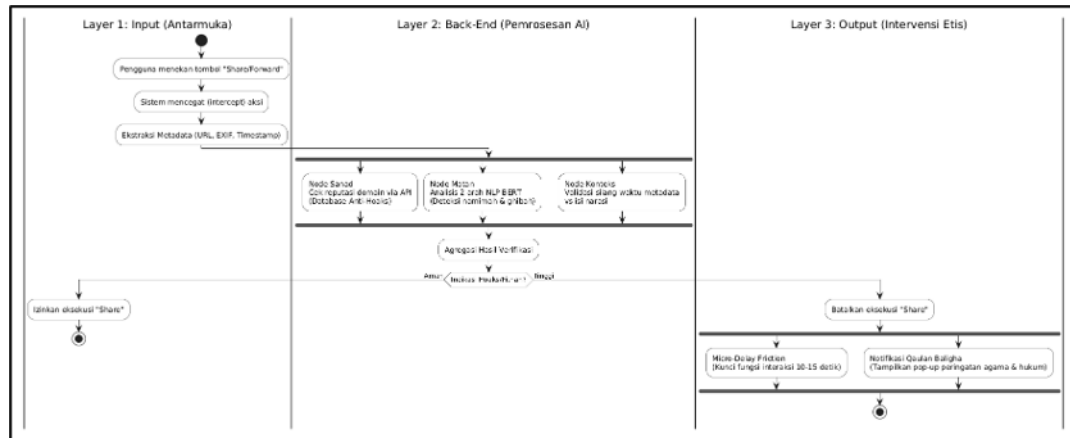
Antarmuka aplikasi melakukan pencegahan (*intercept*) HTTP request tepat saat pengguna menekan tombol *Share*, *Forward*, atau *Retweet*. Secara senyap (*background process*), sistem mengekstraksi *payload* berupa teks narasi, URL, dan tanda tangan file media, lalu mem-versikannya ke server pusat tanpa membekukan layar (*freeze*) secara prematur.

2. Layer 2: Core Processing Engine (*Server-Side*)

Lapisan ini melakukan agregasi konsensus. *Node Sanad*, *Node Matan*, dan *Node Konteks* memberikan skor independen. Sistem mengalkulasi bobot keputusan (*weighted scoring*) di mana integritas Matan dan Konteks diberikan bobot penalti tertinggi. Jika nilai probabilitas kepalsuan agregat melampaui ambang batas (*threshold*) kritis 75%, sebuah token *flagging* diaktifkan dan dikirim kembali ke klien.

3. Layer 3: Output & User Intervention (*User Interface Modification*)

Mengubah struktur *user interface* dari pasif menjadi dinamis-represif. Perintah jaringan memblokir transmisi post ke publik dan memicu eksekusi *Micro-Delay Friction* berupa penguncian ketukan layar (*touch event listener disabled*) yang diiringi dengan injeksi *pop-up* dialog. Untuk memenuhi target NFR latensi yang amat singkat tersebut di tengah tingginya arus transaksi data massal (*high-throughput real-time intercept*), arsitektur *Microservices* ini dirancang dengan mekanisme *caching* berlapis pada *edge server* dan strategi *auto-scaling* untuk menghindari hambatan jalur data (*bottleneck*). Terkait jaminan privasi data (*data privacy*) saat proses pencegahan (*intercept*), sistem mengimplementasikan protokol enkripsi asimetris secara *end-to-end* (E2EE), di mana algoritma hanya diizinkan untuk mengekstrak metadata dan nilai *hash* tanpa merekam maupun menyimpan profil data pribadi pengguna secara permanen di server. Untuk memvisualisasikan sinkronisasi antar-lapisan ini dari awal pemicuan hingga eksekusi sanksi antarmuka, penulis merancang pemodelan ke dalam bentuk *Activity Diagram* (lihat Gambar 1).



Gambar 1. Activity diagram alur kerja sistem verifikasi fakta berbasis nilai *tabayyun*

Gambar 1 mengilustrasikan arsitektur sistem verifikasi fakta dalam bentuk UML *Activity Diagram* yang terbagi ke dalam tiga area kerja (*swimlane*) utama. Proses diinisiasi pada Layer 1: Input (Antarmuka) ketika pengguna menekan tombol "Share/Forward". Sistem secara proaktif mencegah (*intercept*) aksi tersebut dan mengekstraksi metadata dasar (URL, EXIF, dan *Timestamp*). Alur data kemudian bergerak menuju Layer 2: Back-End (Pemrosesan AI), di mana sebuah *fork node* memecah proses evaluasi menjadi tiga komputasi paralel: *Node Sanad* (verifikasi reputasi domain via API Database Anti-Hoaks), *Node Matan* (analisis dua arah NLP BERT untuk mendeteksi *namimah* dan *ghibah*), serta *Node Konteks* (validasi silang metadata waktu terhadap narasi).

Ketiga hasil proses komputasi paralel ini disatukan kembali pada tahapan "Agregasi Hasil Verifikasi". Selanjutnya, sebuah *decision node* bertugas mengevaluasi tingkat ancaman; jika indikasi hoaks/fitnah bernilai "Aman", sistem mengembalikan alur ke Layer 1 untuk mengizinkan eksekusi transmisi pesan "Share". Namun, jika terdeteksi tingkat ancaman "Tinggi", alur akan dialihkan ke Layer 3: Output (Intervensi Etis). Pada lapisan ini, eksekusi penyebaran pesan secara mekanis dibatalkan, diikuti oleh penegakan sanksi antarmuka secara bersamaan: aktivasi *Micro-Delay Friction* (mengunci interaksi layar selama 10-15 detik) dan pemunculan "Notifikasi *Qaulan Baligha*" (peringatan agama dan hukum). Pemodelan visual ini secara komprehensif merepresentasikan transformasi prinsip *tabayyun* menjadi mekanisme penjaga gawang (*gatekeeper*) preventif pada level arsitektur sistem.

Analisis Sistem berdasarkan Anatomi Disinformasi AI

Rancangan arsitektur perangkat lunak yang telah dipaparkan pada bagian hasil di atas merupakan sebuah respons metodologis terhadap eskalasi ancaman AI generatif. Lanskap siber saat ini didominasi oleh algoritma *Generative Adversarial Networks* (GANs) yang memproduksi *deepfake* hiper-realistis, di mana manipulasi audio-visual direkayasa dengan tingkat kemiripan tak kasat mata. Para mufassir dan cendekiawan kontemporer menilai fenomena fabrikasi massal tanpa moralitas ini bukan sekadar kejahatan siber konvensional, melainkan wujud penyimpangan komunikasi tingkat tinggi (kategori *syubhat* dan fitnah *kubra*) yang secara sistematis menargetkan hancurnya integrasi nasional dan akidah umat (Mukti et al., 2025).

Untuk merespons krisis epistemologis ini, perintah *tabayyun* dalam Al-Qur'an harus didudukkan kembali dan diekspansi maknanya. Merujuk pada Tafsir Al-Azhar karya

Buya Hamka, Surah Al-Hujurat ayat 6 diturunkan (*asbabun nuzul*) sebagai peringatan keras agar umat tidak menelan mentah-mentah berita dari "orang fasik" (Hamka, 1982). Tafsir tersebut menegaskan bahwa keengganan memverifikasi informasi akan menyeret peradaban pada kebutaan rasionalitas (*jahalah*) dan penyesalan massal akibat tindakan zalim. Dalam konteks disinformasi generatif hari ini, definisi "fasik" mengalami evolusi entitas; ia tidak lagi hanya berwujud manusia fisik, melainkan telah bermutasi menjadi mesin komputasi—berupa kecerdasan algoritma, perangkat lunak manipulatif, dan jaringan bot nir-identitas yang bekerja eksponensial tanpa batas kelelahan moral.

Menghadapi kecepatan mesin fasik ini, mengandalkan "*tabayyun* kognitif"—yakni sekadar imbauan normatif agar audiens mengecek fakta secara mandiri—telah terbukti gagal total di lapangan (Latifah, 2020). Berdasarkan teori psikologi kognitif Daniel Kahneman, manusia memiliki pemikiran cepat reaktif (*System 1*) dan pemikiran lambat analitis (*System 2*). Arsitektur antarmuka media sosial dirancang secara manipulatif untuk mengurung pengguna dalam *echo chamber* (ruang gema) yang secara eksklusif hanya menstimulasi *System 1*. Algoritma secara agresif menyuplai informasi provokatif guna merangsang lonjakan hormon dopamin dan kortisol di amigdala (pusat emosi). Akibatnya, korteks prefrontal (pusat logika) lumpuh seketika. Pengguna terstimulasi untuk menjadi impulsif, menekan tombol *share* sebelum *System 2* sempat mengambil alih untuk memproses *tabayyun* (Noor, 2018). Oleh karena itu, *System Requirements* komputasional yang digagas penulis dilandaskan pada kaidah ushul fiqih *Sadd al-Dzari'ah* (menutup jalan menuju kerusakan). Otoritas verifikasi harus dipindahkan dari pundak pengguna ke dalam arsitektur algoritma server sebagai penjaga gerbang yang tak bisa diintervensi oleh emosi sesaat.

Kesesuaian Fiqih Informasi dan Etika *Qaulan Baligha*

Modul tiga lapis pada *Layer 2* yang mengolaborasikan agregasi NLP dan *Perceptual Hashing* merupakan terjemahan komputasional yang paling presisi dari aturan normatif Fatwa MUI No. 24 Tahun 2017. Fatwa tersebut memberikan panduan syar'i untuk mengidentifikasi hoaks dengan memeriksa tiga ranah: keabsahan sumber riwayat (*sanad*), kebenaran materi teks (*matan*), dan akurasi ruang waktu (konteks) (Syarif & Kurniati, 2025). Penggunaan mesin BERT pada rancangan arsitektur ini mengadopsi ketelitian disiplin ilmu *Jarh wa Ta'dil*. Mesin tidak hanya membaca huruf, melainkan mengekstrak niat tersembunyi di balik kalimat (*semantic intent*) untuk mencegah terjadinya penyebaran namimah (adu domba) dan *hate speech* berbasis SARA sebagaimana diatur ketat dalam konsensus Fiqih Informasi (Novianti, 2023).

Selain ketangguhan *backend*, kritik etika terkuat dari penelitian ini ditujukan pada filosofi desain aplikasi antarmuka (*front-end*). Paradigma *User Experience* (UX) komersial saat ini terlalu mendewakan prinsip *frictionless design* (desain tanpa hambatan), seperti fitur *one-click share* dan *infinite scroll*. Dalam kacamata ilmu komputer, ini dikenal sebagai *Dark Patterns*—desain yang secara sengaja mengeksploitasi dan memfasilitasi sifat '*ajalah* (ketergesaan) penggunaanya demi menciptakan *traffic* data yang menguntungkan korporat. Secara teologis, desain yang secara sistematis melumpuhkan kewajiban *ihtiyath* (kehati-hatian) seorang muslim ini dianggap menyalahi pilar *Maqashid Syariah*, khususnya *Hifdzul Aql* (perlindungan terhadap fungsi nalar dan akal sehat).

Sebagai solusi teknis, rancangan *Micro-Delay Friction* (penguncian tombol 10-15 detik) di *Layer 3* diperkenalkan sebagai *Anti-Dark Pattern* (hambatan etis). Secara neurosains, jeda paksa yang tak bisa dilewati (*unskippable interruption*) ini memaksa otak

mereda, memutus siklus dopamin yang meledak saat pengguna merasa marah atau takjub, sehingga memberikan waktu yang cukup bagi rasionalitas (*System 2*) untuk mengambil kendali ulang atas emosi primitif. Manuver *value-sensitive design* ini adalah wujud nyata materialisasi etika komunikasi Islam ke dalam layar (Efendi et al., 2024; Jalil & Rachman, 2020). Selain itu, desain *Choice Architecture* (arsitektur pilihan) pada *pop-up* dialog membuang bahasa mesin *error* dan menggantinya dengan prinsip linguistik *Qaulan Baligha* (pesan keagamaan yang lugas, transparan, dan menyentuh sanubari). Layar akan menampilkan rincian hukum negara bersanding dengan peringatan dosa jariyah fitnah. Kerangka *ethical-by-design* ini menjadi instrumen *nudging* moral yang sangat mutakhir untuk mendisiplinkan tabiat komunikasi dan merekayasa ulang literasi digital audiens secara sistematis, massal, dan preventif.

KESIMPULAN

Berdasarkan sintesis interdisipliner yang telah dilakukan, penelitian ini menyimpulkan bahwa Fatwa MUI No. 24 Tahun 2017 dan konsep Fiqih Informasi memiliki daya aplikatif yang sangat kuat untuk dibedah dan dirakit ulang menjadi *System Requirements Specification* (SRS) dalam kerangka arsitektur perangkat lunak. Masalah disinformasi generatif AI tidak dapat lagi diselesaikan semata-mata dengan literasi kognitif pasif, melainkan membutuhkan intervensi teknis yang berakar pada nilai-nilai etika lokal. Penelitian ini membuktikan secara konseptual bahwa penggabungan kecerdasan algoritma *back-end* yakni pelacakan rekam jejak (*sanad*), analisis NLP BERT untuk integritas teks (*matan*), dan ekstraksi telemetri (konteks) yang dikawinkan secara paralel dengan intervensi *front-end* berupa *interface friction* (jeda waktu interaksi) terbukti mampu menjadi solusi pencegahan disinformasi yang komprehensif. Desain *ethical-by-design* ini tidak hanya mendeteksi kebohongan secara komputasional, tetapi juga mampu secara mekanis merekayasa ulang tabiat komunikasi pengguna dan menyabotase sifat tergesa-gesa (*'ajalah*), sehingga prinsip *tabayyun* dapat ditegakkan secara otomatis dan sistematis.

Sebagai sebuah perancangan konseptual, penelitian ini tentu memiliki keterbatasan karena belum diimplementasikan secara empiris. Oleh karena itu, saran untuk penelitian dan pengembangan selanjutnya adalah urgensi untuk mentransformasikan model arsitektur teoretis ini ke dalam purwarupa (*prototype*) aplikasi yang berfungsi penuh di lingkungan laboratorium. Validasi empiris dan uji keandalan wajib dilakukan di masa depan dengan mengukur tingkat kebergunaan antarmuka menggunakan instrumen *System Usability Scale* (SUS), serta mengevaluasi performa algoritma deteksi BERT secara kuantitatif melalui pengujian *F1-Score*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Sains dan Teknologi, Universitas Islam Negeri Alauddin Makassar yang telah memberikan fasilitas, literatur, dan dukungan akademik sehingga gagasan penelitian rekayasa perangkat lunak antardisiplin ini dapat diselesaikan.

DAFTAR PUSTAKA

- Cox, A. L., Gould, S., Cecchinato, M. E., Iacovides, I., & Renfree, I. (2016). Design frictions for mindful interactions: The case for microboundaries. *Conference on Human Factors in Computing Systems - Proceedings*, 1389–1397. <https://doi.org/10.1145/2851581.2892410>.
- Efendi, E., Salsabila, A. Z., Jannah, S. W. N., & Hidayat, W. (2024). Konsep Informasi dalam Perspektif Islam. *INNOVATIVE: Journal Of Social Science Research*, 3(2), 31–38.

- <https://doi.org/10.35719/maddah.v3i1.14>.
- Fardhina, A., Siregar, R. M., Br Sibarani, M. R. W., Br Ginting, I. C., & Pratama, A. (2025). Sistem deteksi berita hoaks berbasis algoritma *natural language processing* (NLP) menggunakan BERT. *Jurnal Manajemen Informatika, Sistem Informasi dan Teknologi Komputer (JUMISTIK)*, 4(1), 450–461. <https://doi.org/10.70247/jumistik.v4i1.156>.
- Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. *Conference on Human Factors in Computing Systems - Proceedings*, 534, 1–14. <https://doi.org/10.1145/3173574.3174108>.
- Hamka, B. (1982). *Tafsir Al-Azhar Jilid 9 (Juzu' 25)*. Jakarta: Pustaka Panjimas.
- Hastuti, Asmia, & Ruhaya, B. (2023). Konsolidasi fatwa MUI dengan fikih informasi dalam merumuskan etika bermuamalah di media sosial. *Risalah: Jurnal Pendidikan dan Studi Islam*, 9(2), 731–746. https://doi.org/10.31943/jurnal_risalah.v9i2.371.
- Jalil, M., & Rachman, E. S. (2020). Etika komunikasi Islam dalam kegiatan pembelajaran online. *Ath Thariq Jurnal Dakwah dan Komunikasi*, 4(2), 192–208. https://doi.org/10.32332/ath_thariq.v4i2.2619.
- Latifah, E. (2020). Efektifitas tabayyun di media online bagi generasi milenial. *Alamtara: Jurnal Komunikasi Dan Penyiaran Islam*, 4(1), 18–25.
- Lembaga Fatwa Majelis Ulama Indonesia. (2024). *Fatwa Majelis Ulama Indonesia Nomor 24 Tahun 2017 Hukum dan pedoman bermuamalah melalui media sosial*. Majelis Ulama Indonesia, 1. <https://www.fatwamui.com>.
- Mastori, Bahtiar, A. T., Gunawan, C., & Dale, R. (2025). Digital islamic propagation communication strategies in countering religious disinformation in the post-truth era: A literature study. *Komunida: Media Komunikasi dan Dakwah*, 15(2), 98–120. <https://doi.org/10.35905/komunida>.
- Mukti, A., Fathoni, A. N., Hakim, M. L., Cahyono, A. M., Bahrudin, M., & Widiyaningrum, K. M. (2025). Ayat-Ayat tabayyun dan isu hoaks: kajian tafsir al-qur'an dalam konteks literasi digital modern. *Jurnal Multidisiplin West Science*, 2(01), 155–166. <https://doi.org/10.58812/jmws.v2i01.3060>.
- Ngarifin, M. K., & Sutopo. (2025). Fatwa Majelis Ulama Indonesia (MUI) tentang berpedoman dalam bermedia sosial (upaya mencegah konten media sosial yang berisi berita bohong dan penyebaran permusuhan). *The Republic: Journal of Constitutional Law*, 3(1), 27–33. <https://doi.org/10.55352/htn.v3i1.1707>.
- Noor, M. U. (2018). Penilaian kualitas informasi sebagai bentuk sikap tabayyun ketika menerima informasi di sosial media dan internet. *BIBLIOTIKA: Jurnal Kajian Perpustakaan dan Informasi Volume*, 2(1), 33–40. <https://doi.org/10.17977/um008v2i12018p033>.
- Novianti, N. F. (2023). Relevansi etika komunikasi islam dalam bermedia sosial dengan fatwa MUI Nomor 24 Tahun 2017. *Journal of Islamic Sosial Science and Communication*, 2(2), 87–98. <https://doi.org/10.54801/jisscdiksi.v2i2.199>.
- Nurdin, S. W., & Nugraha, I. F. (2025). Ancaman deepfake dan disinformasi berbasis AI: Implikasi terhadap keamanan siber dan stabilitas nasional Indonesia. *JIMR: Journal Of International Multidisciplinary Research*, 4(1), 73–92. <https://doi.org/10.62668/jimr.v4i01.1551>.
- Syarif, M. I., & Kurniati. (2025). Mediasi keagamaan MUI dalam fatwa Nomor 24 Tahun 2017 dan analisis Maqāṣid Al-Syarī'Ah di era digital. *Bustanul Fuqaha: Jurnal Bidang Hukum Islam*, 6(3), 684–707. <https://doi.org/10.36701/bustanul.v6i3.2719>.
- Zaini, M., Kusmawati, W., Sari, R. S., & Ernata, Y. (2026). Etika digital & akhlak dalam islam : studi fenomenologis media sosial di Indonesia. *Paradigma: Jurnal Filsafat, Sains, Teknologi, dan Sosial Budaya*, 32(1), 60–71. <https://doi.org/10.33503/paradigma.v32i1.2873>.