

## Revolusi Ilmiah dalam Pengembangan *Machine Learning* untuk Penjaminan Kualitas Perangkat Lunak

Nyoman Ayu Gita Gayatri

Doctoral Computer Science, Bina Nusantara University, Indonesia  
nyoman.gayatri@binus.ac.id

---

### ABSTRACT

*The integration of Machine Learning (ML) into Software Quality Assurance (SQA) is commonly discussed as a technical advancement in defect prediction, test automation, and software reliability. Such a perspective, however, is insufficient for understanding the human consequences of this transformation. This study repositions ML-driven SQA as a scientific, ethical, and character-development issue in the transition towards Industry 5.0. Drawing on Thomas S. Kuhn's theory of scientific revolutions, Abraham Maslow's psychology of science, and the principles of Trustworthy AI, this article examines how the transition from deterministic testing to learning-based quality assurance changes the professional identity and moral responsibilities of software engineers. The study employs a qualitative interdisciplinary design through a Systematic Literature Review, Conceptual Framework Analysis, and philosophical hermeneutics. The synthesis indicates that conventional SQA has entered a Kuhnian crisis because rule-based testing is increasingly unable to accommodate the uncertainty, scale, and dynamic interdependence of cloud-native and AI-enabled systems. ML introduces a new, probabilistic paradigm that is incommensurable with the former paradigm in terms of evidence, decision-making, and the role of engineers. The article argues that this transition requires engineers to move beyond a safety-oriented mentality characterized by dependence on certainty and procedural control toward a growth-oriented character marked by epistemic courage, adaptability, humility, ethical responsibility, and human-AI collaboration. The study proposes the Kuhn-Maslow-Trustworthy AI Framework as a conceptual foundation for defining the character of engineers in Industry 5.0. The findings emphasize that trustworthy and explainable AI should not be treated as optional technical features, but as moral infrastructures for responsible software quality decisions.*

**Keywords:** *Character development, machine learning, software quality assurance, trustworthy AI*

### ABSTRAK

*Pemanfaatan Machine Learning (ML) dalam penjaminan kualitas perangkat lunak atau Software Quality Assurance (SQA) umumnya dipahami sebagai kemajuan teknis dalam prediksi cacat, otomatisasi pengujian, dan peningkatan reliabilitas sistem. Namun, pandangan yang semata-mata teknis tidak cukup untuk menjelaskan konsekuensi manusiawi dari transformasi tersebut. Penelitian ini memosisikan SQA berbasis ML sebagai persoalan revolusi ilmiah, etika teknologi, dan pengembangan karakter dalam transisi menuju Industri 5.0. Dengan menggunakan teori revolusi ilmiah Thomas S. Kuhn, psikologi sains Abraham Maslow, serta prinsip Trustworthy AI, artikel ini menganalisis bagaimana peralihan dari pengujian deterministik menuju penjaminan kualitas berbasis pembelajaran mengubah identitas profesional dan tanggung jawab moral rekayasawan perangkat lunak. Penelitian menggunakan desain kualitatif interdisipliner melalui Systematic Literature Review, Conceptual Framework Analysis, dan hermeneutika filosofis. Hasil sintesis menunjukkan bahwa SQA konvensional telah memasuki krisis Kuhnian karena pengujian berbasis aturan semakin tidak mampu menghadapi ketidakpastian, skala, dan keterhubungan dinamis dalam sistem cloud-native serta sistem berbasis AI. Machine Learning menghadirkan paradigma probabilistik yang tidak sepenuhnya sepadan dengan*

*paradigma sebelumnya dalam hal bukti, pengambilan keputusan, dan peran manusia. Artikel ini berargumen bahwa transformasi tersebut menuntut rekayasawan melampaui mentalitas keamanan yang terlalu bergantung pada kepastian prosedural menuju karakter pertumbuhan yang ditandai oleh keberanian epistemik, adaptabilitas, kerendahan hati intelektual, tanggung jawab etis, dan kemampuan berkolaborasi dengan AI. Penelitian ini mengusulkan Kerangka Kuhn–Maslow–Trustworthy AI sebagai dasar konseptual untuk mendefinisikan karakter rekayasawan era Industri 5.0. Temuan penelitian menegaskan bahwa AI yang dapat dipercaya dan dijelaskan bukan sekadar fitur teknis tambahan, tetapi infrastruktur moral bagi keputusan kualitas perangkat lunak yang bertanggung jawab.*

**Kata Kunci:** Pengembangan karakter, machine learning, software quality assurance, kecerdasan buatan yang terpercaya

## I. PENDAHULUAN

Perangkat lunak kini tidak lagi hanya menjalankan fungsi administratif di belakang layar. Ia menentukan bagaimana layanan kesehatan dijadwalkan, transaksi keuangan diverifikasi, pembelajaran diselenggarakan, dan keputusan organisasi dibuat. Karena itu, kualitas perangkat lunak tidak dapat diperlakukan sebagai urusan teknis yang berdiri sendiri. Kegagalan sistem dapat berakibat pada hilangnya kepercayaan publik, gangguan layanan, kerugian ekonomi, atau keputusan yang tidak adil bagi pengguna. Dalam peralihan menuju Industry 5.0, teknologi digital justru diharapkan semakin berpusat pada manusia (*human-centric*), berkelanjutan, dan resilien; manusia tidak disingkirkan dari proses rekayasa, melainkan ditempatkan sebagai pihak yang bertanggung jawab atas arah penggunaan teknologi (Yao et al., 2024; UNESCO, 2021).

Tuntutan tersebut berhadapan dengan perubahan nyata pada lingkungan pengembangan perangkat lunak. Sistem *cloud-native*, *microservices*, *containerization*, dan *Continuous Integration/Continuous Deployment (CI/CD)* mempercepat pengiriman fitur, tetapi juga memperbanyak interaksi antar-komponen, variasi konfigurasi, serta sumber ketidakpastian. Dalam kondisi demikian, pengujian deterministik yang mengandalkan skrip tetap, aturan *if-then*, dan pemeriksaan pasca-kode menjadi semakin mahal untuk dipelihara. Kajian tentang seleksi dan prioritasasi *test case* berbasis ML menunjukkan bahwa frekuensi integrasi yang tinggi membuat pelaksanaan seluruh *test suite* tidak selalu realistis; tim perlu memilih tes yang paling informatif dan berpotensi menemukan kegagalan lebih awal (Pan et al., 2022). Masalah *test flakiness* juga memperlihatkan bahwa hasil tes otomatis dapat berubah bukan karena fungsi bisnis benar-benar rusak, melainkan karena variasi waktu, lingkungan, data, atau ketergantungan eksternal (Ngo et al., 2022).

*Machine learning* menawarkan cara kerja yang berbeda. Alih-alih menuliskan seluruh kemungkinan secara eksplisit, model belajar dari data perubahan kode, riwayat kegagalan, log eksekusi, atau telemetri produksi untuk memperkirakan risiko dan memprioritaskan perhatian. Tinjauan sistematis Wang et al. (2023) memperlihatkan bahwa *machine/deep learning* telah digunakan pada berbagai aktivitas rekayasa perangkat lunak, termasuk prediksi cacat, klasifikasi artefak, dan pengambilan keputusan berbasis data. Pada *automation test*, praktik industri juga bergerak ke arah pembuatan tes otomatis, pemeliharaan skrip, dan *self-healing scripts* berbantuan AI untuk menghadapi biaya pemeliharaan yang tinggi (Ricca et al., 2025).

Namun, artikel ini tidak menganggap ML sebagai obat mujarab ataupun bukti bahwa manusia dapat dilepaskan dari penjaminan kualitas. Justru sebaliknya, semakin besar otonomi

analitik yang diberikan kepada sistem, semakin tinggi tuntutan karakter dan tanggung jawab pada orang yang merancang, melatih, memvalidasi, serta menggunakan sistem tersebut. *National Institute of Standards and Technology* menegaskan bahwa AI adalah sistem sosio-teknis: risiko dan manfaatnya muncul dari interaksi model, data, organisasi, pengguna, dan konteks penerapan. Karena itu, AI yang dapat dipercaya harus valid, andal, serta aman terhadap gangguan, akuntabel dan transparan, dapat dijelaskan dan ditafsirkan, menjaga privasi, serta mengelola bias yang merugikan (Tabassi, 2023).

### 1.1 Kesenjangan Penelitian dan Rumusan Masalah

Literatur teknis telah banyak membahas akurasi model, pemilihan fitur, prediksi cacat, dan otomasi pengujian. Literatur etika AI juga telah merumuskan prinsip transparansi, akuntabilitas, keadilan, dan *human oversight*. Akan tetapi, masih terbatas penelitian yang menjembatani tiga pertanyaan secara terpadu. Pertama, apakah peralihan ke *ML-supported SQA* merupakan peningkatan bertahap atau perubahan paradigma? Kedua, mengapa perubahan teknis tersebut sering menimbulkan resistensi, kecemasan profesional, atau keengganan beradaptasi pada rekayasawan? Ketiga, karakter apa yang perlu dikembangkan agar penggunaan AI dalam keputusan kualitas tetap bertanggung jawab secara moral?

Kesenjangan ini penting bagi upaya-upaya pembentukan dan pengembangan karakter karena prosesnya tidak berhenti di ruang kelas atau relasi sosial sehari-hari melainkan juga dibentuk ketika seseorang menghadapi ketidakpastian, menggunakan kekuasaan teknologis, dan memutuskan apakah akan mempercayai atau menolak rekomendasi sistem otomatis. Dalam konteks ini, pengembangan karakter rekayasawan adalah pengembangan kapasitas moral-intelektual untuk bertindak bijaksana di bawah kondisi kompleks, bukan sekadar pembelajaran keterampilan teknis. Gagasan *technomoral virtues* dari Vallor (2016) dan *virtue-based framework* untuk praktik etika AI dari Hagendorff (2022) memberikan dasar bahwa teknologi yang baik membutuhkan pembentukan kebiasaan, penilaian, dan disposisi manusia yang baik.

### 1.2 Tujuan, Kebaruan, dan Kontribusi

Penelitian ini bertujuan membangun sintesis konseptual mengenai revolusi ML dalam SQA dengan menempatkan transformasi karakter rekayasawan sebagai inti pembahasannya. Kebaruan artikel terletak pada penggabungan tiga tradisi yang lazim berjalan sendiri-sendiri: (1) filsafat sains Kuhn untuk membaca perubahan paradigma, (2) psikologi humanistik Maslow untuk menjelaskan tegangan antara orientasi keamanan dan pertumbuhan, serta (3) kecerdasan buatan yang terpercaya (*trustworthy AI*) untuk menerjemahkan nilai etis ke dalam tata kelola rekayasa. Hasilnya adalah model Karakter Rekayasawan Era Industri 5.0 yang menghubungkan perubahan metode kerja dengan keberanian epistemik, kerendahan hati, akuntabilitas, keadilan, kepedulian, dan kolaborasi manusia-AI.

Kontribusi praktisnya adalah pergeseran fokus: organisasi tidak cukup mengukur keberhasilan adopsi AI dari kecepatan rilis, jumlah tes, atau akurasi prediksi. Organisasi perlu menilai apakah rekayasawan mampu menjelaskan batas model, mempertanyakan keputusan yang berisiko, melindungi data, mendokumentasikan alasan keputusan, dan mempertahankan kewenangan manusia pada titik-titik kritis. Dengan demikian, SQA bukan hanya gerbang

teknis untuk merilis produk, melainkan arena pendidikan karakter profesional dalam ekosistem digital.

## II. METODE PENELITIAN

Penelitian ini menggunakan desain kualitatif-interdisipliner dengan tiga lapis analisis: *Systematic Integrative Literature Review (SILR)*, *Conceptual Framework Analysis (CFA)*, dan hermeneutika filosofis. Pemilihan desain ini didasarkan pada tujuan penelitian yang bukan menghitung besar pengaruh satu algoritme, melainkan menjelaskan hubungan antara perubahan teknis, perubahan paradigma, orientasi psikologis, dan tuntutan etis. *Systematic literature review* dalam rekayasa perangkat lunak menuntut strategi pencarian, seleksi, dan sintesis yang transparan agar kesimpulan tidak bertumpu pada contoh yang dipilih secara serampangan (Kitchenham et al., 2009). Sementara itu, CFA tepat digunakan untuk membangun jaringan konsep dari tubuh pengetahuan multidisipliner yang tidak homogen (Jabareen, 2009).

### 2.1 Pertanyaan Kajian

- RQ1: Anomali teknis dan metodologis apa yang membuat SQA tradisional semakin tidak memadai dalam ekosistem *cloud-native* dan CI/CD?
- RQ2: Bagaimana studi tentang *ML-assisted testing*, *defect prediction*, *test prioritization*, dan *test automation* menjelaskan perubahan cara kerja SQA?
- RQ3: Bagaimana konsep *normal science*, *anomaly*, *crisis*, *revolution*, dan *incommensurability* dari Kuhn dapat menerangkan perubahan tersebut?
- RQ4: Bagaimana *safety science* dan *growth science* dari Maslow dapat dipakai secara hati-hati untuk menafsirkan *technical reluctance* pada rekayasawan?
- RQ5: Nilai karakter dan mekanisme *Trustworthy AI* apa yang perlu dikembangkan agar *AI-assisted SQA* mendukung, bukan mengaburkan, tanggung jawab manusia?

### 2.2 Protokol Penelusuran dan Seleksi Literatur

Penelusuran dilakukan terhadap karya fondasional dan penelitian empiris/ulasan yang relevan dengan empat domain: filsafat sains, psikologi sains dan karakter, AI etis, serta *software quality assurance*. Basis data yang menjadi rujukan penelusuran meliputi Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, dan laman resmi lembaga standar. Kata kunci dikombinasikan dengan operator Boolean, misalnya: ("*machine learning*" OR "*artificial intelligence*") AND ("*software testing*" OR "*software quality assurance*" OR "*defect prediction*"); ("*trustworthy AI*" OR "*explainable AI*") AND ("*engineering*" OR "*software*"); serta ("*Kuhn*" OR "*scientific revolution*") AND ("*technology*" OR "*engineering*").

Artikel diikutkan apabila: (a) diterbitkan pada kanal ilmiah atau lembaga standar yang dapat diverifikasi; (b) membahas secara langsung SQA, pengujian, prediksi cacat, otomasi tes, AI ethics, Trustworthy AI, filsafat sains, atau psikologi sains; (c) menyediakan penjelasan metodologis atau kontribusi konseptual yang jelas; dan (d) terbit terutama dalam rentang 2016-2026 untuk literatur teknis, dengan pengecualian terhadap karya klasik Kuhn dan Maslow yang memang menjadi landasan teori dari artikel ini. Artikel yang semata-mata mempromosikan produk, tidak menjelaskan metode, atau tidak memiliki keterkaitan substantif dengan

pertanyaan penelitian dikeluarkan. Prinsip pelaporan seleksi mengacu pada transparansi PRISMA 2020, tetapi kajian ini tidak menyajikan meta-analisis statistik karena desain, konteks, metrik, dan unit analisis studi yang dikaji sangat heterogen (Page et al., 2021).

**Tabel 1. Protokol analisis *Systematic Integrative Literature Review*, CFA, dan hermeneutika**

Tahap	Aktivitas	Keluaran Analitis
Identifikasi	Menelusuri karya fondasional dan studi empiris tentang SQA, ML, AI ethics, Industry 5.0, Kuhn, dan Maslow.	Korpus literatur lintas-disiplin yang dapat ditelusuri sumbernya.
Seleksi	Menerapkan kriteria relevansi, kualitas kanal publikasi, transparansi metode, dan keterhubungan dengan RQ.	Artikel primer dan dokumen standar yang memadai untuk sintesis.
Ekstraksi	Mencatat masalah SQA, teknik ML, bukti penerapan, risiko, posisi manusia, serta konsep karakter/etika.	Matriks konsep: teknis, filosofis, psikologis, dan etis.
CFA	Menghubungkan konsep menjadi model: Kuhn → Maslow → Trustworthy AI → karakter rekayasawan.	Kerangka Karakter Rekayasawan Era Industri 5.0.
Hermeneutika	Menafsirkan <i>technical reluctance</i> sebagai pengalaman makna: kehilangan kontrol, identitas, dan kepastian; bukan sebagai kelemahan personal.	Penjelasan psikologis dan rekomendasi pembelajaran/organisasi.

### 2.3 *Conceptual Framework Analysis* dan Pembacaan Hermeneutika

Pada tahap CFA, setiap sumber dikodekan ke dalam empat kelompok konsep. Kelompok pertama adalah kondisi teknis SQA: kompleksitas, *test flakiness*, biaya pemeliharaan, *test selection*, *defect prediction*, dan *self-healing*. Kelompok kedua adalah konsep Kuhnian: *normal science*, *anomaly*, *crisis*, *revolution*, dan *incommensurability*. Kelompok ketiga adalah orientasi psikologis: *safety*, *growth*, *means-centering*, *problem-centering*, serta keengganan beradaptasi. Kelompok keempat adalah etika dan tata kelola: validitas, reliabilitas, transparansi, *explainability*, *fairness*, privasi, keamanan, dan akuntabilitas. Matriks ini kemudian disintesis untuk menemukan hubungan sebab-akibat, tegangan, dan implikasi karakter.

Hermeneutika digunakan bukan untuk mengklaim diagnosis psikologis terhadap individu rekayasawan, melainkan untuk memahami makna yang mungkin melekat pada resistensi terhadap AI. Dalam tradisi hermeneutika, pemahaman dibangun melalui dialog antara teks, konteks, dan horison penafsir; pembacaan tidak berhenti pada gejala lahiriah, tetapi mencari struktur makna yang mendasarinya (Gadamer, 2004). Karena itu, *technical reluctance* dalam artikel ini diartikan sebagai respons yang dapat dimengerti ketika seorang profesional merasa keahlian lamanya menjadi kurang bernilai, keputusan menjadi lebih sulit diaudit, atau tanggung jawab bergeser kepada sistem yang tidak sepenuhnya dipahami.

Artikel ini tidak menyajikan wawancara primer atau data implementasi dari satu organisasi tertentu. Karena itu, bagian studi kasus disusun sebagai kasus konseptual terapan yang menggabungkan pola empiris dari literatur, bukan sebagai klaim hasil lapangan dari

perusahaan tertentu. Transparansi ini diperlukan agar analisis normatif tidak disamakan sebagai fakta empiris.

## **2.4 Tinjauan Pustaka dan Landasan Konseptual**

### **2.4.1 Revolusi Ilmiah Kuhn: Dari Puzzle-Solving ke Ketidaksepadanan Paradigma**

Kuhn menolak gambaran bahwa ilmu berkembang seperti tumpukan bata: fakta baru ditambahkan sedikit demi sedikit hingga bangunan pengetahuan semakin tinggi. Baginya, perkembangan ilmu bergerak melalui periode sains normal, ketika komunitas bekerja memecahkan *puzzle* dalam seperangkat asumsi bersama, lalu berhadapan dengan anomali yang tidak dapat diselesaikan secara memadai. Jika anomali menumpuk dan kepercayaan pada perangkat lama melemah, krisis membuka ruang bagi revolusi ilmiah dan pembentukan paradigma baru (Kuhn, 2012).

Salah satu konsep penting untuk membaca transisi SQA adalah incommensurability atau ketidaksepadanan. Ketidaksepadanan tidak berarti kedua paradigma tidak dapat dibandingkan sama sekali. Keduanya menggunakan bahasa, standar keberhasilan, dan cara melihat masalah yang tidak sepenuhnya dapat diterjemahkan satu sama lain. Dalam SQA tradisional, keberhasilan lazim dipahami melalui skrip yang lulus/gagal, *coverage*, dan kepatuhan terhadap spesifikasi eksplisit. Dalam SQA berbasis ML, bukti juga mencakup kualitas data, ketidakpastian prediksi, kalibrasi, *drift*, *trade-off false positive/false negative*, dan penjelasan model. Rekayasawan yang tetap menganggap ketidakpastian probabilistik sebagai kegagalan disiplin, misalnya, akan sulit memahami kekuatan dan batas paradigma baru.

Kuhn tidak menyediakan resep untuk memprediksi kapan revolusi terjadi. Namun, kerangkanya berguna sebagai alat diagnosis: ia membantu membedakan antara kesulitan biasa yang masih dapat diatasi dengan perbaikan skrip dan kesulitan yang mengindikasikan keterbatasan cara berpikir. Studi mengenai krisis replikasi di linguistik menunjukkan bagaimana model Kuhn dapat dipakai untuk membaca keresahan metodologis sebagai masa transisi yang menuntut refleksi filosofis dan reformasi praktik, bukan sekadar ketidakberesan administratif (Sönning & Werner, 2021). Analogi ini mendukung pembacaan bahwa konflik seputar AI-assisted SQA bukan semata sengketa alat, melainkan sengketa mengenai apa yang dianggap sebagai pengetahuan yang andal.

### **2.4.2 Maslow, Safety Science, dan Growth Science**

Maslow menaruh perhatian pada manusia yang melakukan sains, bukan hanya pada prosedur sains itu sendiri. Dalam *The Psychology of Science*, ia mengkritik sains yang terlalu berpusat pada cara, instrumen, dan kepastian formal, karena orientasi demikian dapat mengaburkan masalah manusia yang sebenarnya ingin dipecahkan (Maslow, 1966). Pembacaan Kožnjak (2017) atas hubungan Kuhn dan Maslow memperlihatkan bahwa *normal* dan *revolutionary science* dapat dipahami pula melalui tegangan motivasional antara kebutuhan akan keamanan dan kebutuhan akan pertumbuhan.

Dalam artikel ini, *safety science* tidak diartikan sebagai sikap yang buruk atau tidak profesional. Kebutuhan akan keamanan bermanfaat ketika rekayasawan perlu menjaga stabilitas, dokumentasi, dan disiplin prosedural. Masalah muncul apabila keamanan berubah menjadi ketergantungan terhadap kepastian semu: hanya mempercayai apa yang dapat

dikendalikan melalui skrip statis, lalu menolak bukti probabilistik hanya karena tidak menjanjikan kepastian absolut. Sebaliknya, *growth science* tidak identik dengan keberanian mengambil risiko tanpa kendali. Ia berarti kesiapan untuk belajar, mengubah cara kerja, menahan ambiguitas, dan tetap bertanggung jawab ketika pengetahuan belum lengkap.

Pembedaan ini relevan bagi pengembangan karakter. Rekayasawan yang matang tidak dipaksa memilih antara disiplin dan inovasi. Ia mengintegrasikan keduanya: cukup hati-hati untuk tidak menyerahkan keputusan penting secara buta kepada AI, tetapi cukup berani untuk menguji paradigma lama ketika ia tidak lagi memadai. Dengan bahasa etika kebajikan, karakter profesional dibentuk sebagai disposisi untuk menilai secara bijaksana, bukan sekadar menaati daftar aturan (Vallor, 2016; Hagendorff, 2022).

#### **2.4.3 Quality 4.0, Quality 5.0, dan Makna Human-Centric Quality**

Perkembangan manajemen kualitas memperlihatkan pergeseran dari inspeksi reaktif ke pencegahan, pengendalian proses, dan akhirnya kualitas prediktif berbasis data. *Quality 4.0* mengintegrasikan AI, *Internet of Things*, *cloud computing*, dan analitik waktu nyata agar kualitas dapat diamati dan ditingkatkan secara proaktif (Gomaa, 2025). Pada tahap berikutnya, *Quality 5.0* menekankan bahwa kecanggihan digital harus dipandu oleh kolaborasi manusia-mesin, keberlanjutan, keputusan etis, dan perhatian terhadap dampak sosial (Avşar, 2025; Yao et al., 2024).

Implikasinya bagi SQA sangat penting. Kualitas tidak semata angka kegagalan tes atau kecepatan *deployment*. Kualitas mencakup apakah organisasi dapat menjelaskan mengapa suatu sistem dianggap aman untuk digunakan; apakah pengguna dilindungi dari keputusan yang bias; apakah data pengujian diperlakukan secara layak; dan apakah tersedia pihak manusia yang dapat dimintai pertanggungjawaban. Dengan kata lain, kualitas pada era *Industry 5.0* mengandung kualitas karakter dan kualitas tata kelola, bukan hanya kualitas kode.

#### **2.4.4 Machine Learning dalam SQA: Manfaat Teknis dan Batas Translasinya**

*Machine learning* telah digunakan untuk memprediksi modul yang berisiko cacat, memilih dan memprioritaskan *test case*, mengelompokkan kegagalan, mendeteksi anomali, dan membantu pemeliharaan *automation suite*. Pan et al. (2022) meninjau 29 studi primer mengenai *ML-based test case selection* dan *prioritization*, memperlihatkan bahwa informasi parsial tentang kode, tes, dan riwayat eksekusi dapat digabungkan untuk menghasilkan prioritas yang lebih informatif. Pham et al. (2022) menemukan bahwa AI telah memengaruhi berbagai aktivitas *end-to-end test automation*, walaupun manfaat yang dijanjikan tetap bergantung pada kualitas alat, data, dan integrasi proses kerja.

Pada prediksi cacat, *deep learning* dapat mengurangi ketergantungan pada *feature engineering manual* karena model mampu membangun representasi dari data yang kompleks. Namun, keunggulan ini tidak boleh digeneralisasi tanpa syarat. Batool dan Khan (2023) menunjukkan bahwa pendekatan *deep learning* menawarkan potensi performa yang kuat, tetapi juga memperbesar tuntutan terhadap data, komputasi, dan interpretabilitas. Karena itu, pernyataan bahwa ML selalu lebih baik daripada metode tradisional tidak dapat dipertahankan. Pilihan metode harus mempertimbangkan konteks sistem, risiko salah prediksi, kebutuhan audit, dan kemampuan tim untuk mengelolanya.

Referensi tentang *Convolutional Neural Networks (CNN)* dalam deteksi cacat industri tetap relevan sebagai jembatan konseptual, bukan sebagai bukti langsung untuk *defect prediction* pada *source code*. Tinjauan Khanam et al. (2024) menunjukkan bahwa kekuatan CNN berada pada pembelajaran representasi dari data visual yang kompleks untuk inspeksi kualitas. Dalam SQA, analoginya terletak pada prinsip pembelajaran fitur dan deteksi pola anomali; tetapi domain, data, dan risiko keputusan berbeda. Karena itu, artikel ini tidak menggunakan literatur CNN industri untuk menyimpulkan efektivitas SQA secara langsung, melainkan untuk menjelaskan mengapa otomatisasi kualitas berbasis pembelajaran menuntut evaluasi konteks dan tata kelola yang ketat.

#### ***2.4.5 Trustworthy dan Explainable AI sebagai Etika Operasional***

Etika AI tidak boleh berhenti sebagai kumpulan slogan seperti *fairness*, *transparency*, atau *privacy* yang tidak mengubah praktik kerja. Kritik terhadap *AI ethics* menunjukkan bahwa banyak prinsip gagal memberi dampak ketika tidak diterjemahkan menjadi proses, kompetensi, insentif, dan mekanisme evaluasi yang nyata (Hagendorff, 2023). Dalam SQA, risiko tersebut sangat konkret: model dapat memprioritaskan tes secara keliru, mengabaikan kelompok pengguna tertentu, menilai kode aman berdasarkan data historis yang tidak representatif, atau menyembunyikan alasan rekomendasi di balik *black box*.

NIST AI RMF 1.0 menyediakan landasan operasional yang kuat dengan karakteristik *valid* dan *reliable*; *safe*; *secure and resilient*; *accountable and transparent*; *explainable and interpretable*; *privacy-enhanced*; serta *fair with harmful biases managed* (Tabassi, 2023). Prinsip UNESCO juga menekankan martabat manusia, transparansi, *fairness*, keberlanjutan, dan *human oversight* sebagai fondasi etika AI (UNESCO, 2021). Dalam artikel ini, *Explainable AI* bukan sekadar fasilitas visualisasi teknis. XAI adalah syarat agar rekayasawan dapat memeriksa alasan model, mendebat rekomendasi yang meragukan, dan memberi pertanggungjawaban yang dapat dipahami oleh pihak terdampak.

### **III. HASIL DAN PEMBAHASAN**

#### **3.1 Pemetaan Kuhnian atas Evolusi SQA**

Hasil sintesis menunjukkan bahwa perubahan SQA dapat dibaca melalui empat tahap Kuhnian. Pertama, pada fase *normal science*, SQA didominasi pengujian manual dan otomasi deterministik. Tujuan utama tim adalah menerjemahkan kebutuhan menjadi *test case*, menjalankan tes, dan menelusuri penyebab kegagalan. Tahap ini efektif untuk sistem dengan batas yang relatif stabil; keberhasilannya ditopang oleh keyakinan bahwa perilaku sistem dapat didefinisikan secara cukup lengkap melalui aturan dan skenario eksplisit.

Kedua, anomali muncul ketika kecepatan dan kompleksitas ekosistem digital melampaui daya tangkap prosedur lama. Contohnya adalah *test flakiness*, pemeliharaan skrip yang mahal, variasi konfigurasi cloud, serta batas waktu *regression testing* pada *pipeline CI/CD*. Anomali bukan sekadar *bug* tambahan yang dapat diselesaikan dengan menambah jumlah skrip. Ia mempertanyakan asumsi dasar bahwa seluruh perilaku relevan dapat dipetakan terlebih dahulu oleh manusia. Studi Ricca et al. (2025) menunjukkan bahwa pengembangan dan pemeliharaan kode tes merupakan tantangan dominan dalam praktik otomasi, sedangkan

*self-healing scripts* dan *automated test generation* menjadi salah satu respons AI yang paling menonjol.

Ketiga, krisis terjadi ketika organisasi mulai kehilangan kepercayaan pada mekanisme kualitasnya sendiri. *Test suite* yang terlalu lambat menghambat rilis; skrip yang sering gagal palsu menghasilkan kelelahan investigasi; dan cakupan tes tidak lagi menjamin ketahanan layanan nyata. Krisis ini tidak berarti semua metode tradisional harus dibuang. Ia berarti metode lama tidak lagi cukup sebagai pusat pengetahuan. Organisasi membutuhkan bentuk bukti baru: prediksi risiko, pola anomali, probabilitas kegagalan, dan indikator drift.

Keempat, revolusi ditandai oleh hadirnya *ML-supported SQA*. Model menggeser fokus dari pemeriksaan seluruh kemungkinan ke pembelajaran tentang kemungkinan yang paling bermakna. *Test prioritization* menggunakan riwayat perubahan dan kegagalan untuk mempercepat umpan balik; *defect prediction* mengarahkan perhatian ke bagian kode yang lebih berisiko; *anomaly detection* memantau pola yang tidak normal; dan *self-healing* membantu mengurangi kerapuhan tertentu dalam tes antarmuka. Revolusi ini bersifat metodologis sekaligus epistemologis karena pengetahuan kualitas tidak hanya dihasilkan dari aturan yang ditulis, tetapi juga dari inferensi yang dipelajari dari data.

**Tabel 2. Pemetaan paradigma SQA dan tuntutan karakter rekayasawan**

Dimensi	SQA Tradisional	SQA Berbasis ML	Implikasi Karakter
Bentuk pengetahuan	Aturan eksplisit, test case statis, lulus/gagal.	Pola data, probabilitas, <i>confidence</i> , dan deteksi drift.	Kerendahan hati epistemik: mengakui bahwa pengetahuan kualitas selalu parsial.
Orientasi waktu	Reaktif: menemukan cacat setelah atau saat eksekusi.	Prediktif: mengestimasi risiko dan memprioritaskan perhatian lebih awal.	Keberanian epistemik: bertindak berdasarkan bukti yang belum sempurna tetapi terukur.
Peran manusia	Penulis dan pemelihara skrip; <i>gatekeeper</i> .	Perancang konteks, evaluator model, pengawas keputusan, dan penanggung jawab.	Adaptabilitas reflektif dan kesiapan belajar ulang.
Risiko utama	Cakupan tidak memadai, <i>script brittleness</i> , keterlambatan umpan balik.	Bias data, <i>opacity</i> , <i>drift</i> , <i>overreliance</i> , dan privasi.	Integritas, akuntabilitas, dan kewaspadaan kritis.
Standar kualitas	Kepatuhan terhadap spesifikasi dan keberhasilan tes.	Kinerja teknis sekaligus validitas, <i>fairness</i> , <i>explainability</i> , <i>privacy</i> , dan <i>human oversight</i> .	Keadilan, kepedulian, dan kolaborasi manusia-AI.

### 3.2 Ketidaktepatan Paradigma: Mengapa Pelatihan Alat Saja Tidak Cukup

Ketidaktepatan antara SQA tradisional dan SQA berbasis ML terlihat pada empat lapisan. Pertama, bahasa kerja berubah. Istilah *test case*, *assertion*, dan *coverage* tetap penting,

tetapi kini berdampingan dengan dataset, *label noise*, *feature drift*, *calibration*, *false discovery*, dan *model explanation*. Kedua, standar bukti berubah. Laporan pass/fail tidak lagi memadai untuk menjelaskan keputusan model yang memprioritaskan risiko. Ketiga, ukuran keberhasilan berubah. Akurasi yang tinggi dapat menyesatkan ketika model gagal mendeteksi kasus yang jarang tetapi berdampak besar. Keempat, identitas profesional berubah: *QA engineer* tidak cukup menjadi pemelihara skrip, tetapi perlu menjadi penguji asumsi, kurator data, penafsir hasil, dan penjaga batas etis.

Di sinilah konsep *incommensurability* berguna secara kritis. Rekayasawan tidak gagal beradaptasi karena kurang modern atau anti-teknologi. Mereka mungkin sedang menggunakan ukuran keberhasilan yang dibentuk oleh paradigma lama. Seorang penguji yang bertahun-tahun berhasil menjaga stabilitas melalui skrip deterministik dapat memandang output probabilistik sebagai ancaman terhadap kontrol profesional. Respons itu dapat dimengerti. Namun, apabila organisasi hanya menyuruhnya ‘belajar alat AI’, tanpa membahas perubahan standar bukti dan tanggung jawab, pelatihan akan menjadi dangkal. Organisasi perlu menyediakan ruang untuk menafsirkan ulang pengalaman profesional, bukan sekadar memaksakan migrasi teknologi.

### 3.3 Studi Kasus Konseptual Terapan: Tim SQA pada Layanan *Cloud-Native*

Untuk menjawab kritik bahwa argumentasi etis sering terlalu normatif, bagian ini menyajikan studi kasus konseptual terapan. Kasus ini bukan laporan dari perusahaan tertentu. Ia adalah vignette analitis yang dirancang dari pola empiris pada literatur mengenai *test flakiness*, *ML-based test prioritization*, dan *AI-assisted automation* (Ngo et al., 2022; Pan et al., 2022; Ricca et al., 2025).

Bayangkan sebuah perusahaan penyedia layanan digital yang menggunakan arsitektur *microservices* dan merilis pembaruan beberapa kali dalam satu hari. Tim SQA memiliki ratusan skrip regresi antarmuka dan integrasi. Pada awalnya, tim menganggap semakin banyak skrip berarti semakin tinggi kualitas. Dalam praktiknya, sebagian skrip gagal karena perubahan tampilan kecil, latensi layanan pihak ketiga, atau sinkronisasi data yang belum selesai. Penguji menghabiskan banyak waktu untuk mengulang tes dan membedakan kegagalan nyata dari kegagalan palsu. Mereka merasa kewalahan, tetapi juga enggan mempercayai rekomendasi model karena tidak memahami dasar keputusannya.

Organisasi kemudian mengadopsi dua kemampuan berbantuan ML: prioritas tes berdasarkan riwayat perubahan dan kegagalan, serta pemulihan lokator elemen antarmuka yang disertai catatan perubahan. Adopsi tidak langsung berhasil. Sebagian penguji mengkhawatirkan pengurangan peran; sebagian pengembang menganggap model dapat menggantikan validasi manusia; dan manajer tergoda menjadikan skor risiko sebagai satu-satunya dasar keputusan rilis. Pada titik ini, persoalan utama bukan lagi akurasi model, melainkan desain relasi manusia-AI.

**Tabel 3. Studi kasus konseptual: masalah, respons, dan pembelajaran karakter**

Tahap	Masalah dan Respons Teknis	Dinamika Manusia	Pembelajaran Karakter dan Tata Kelola
Sebelum AI	Tes regresi lambat, flaky, dan mahal dipelihara.	Penguji mengandalkan skrip	Menghargai disiplin prosedural tanpa

		karena skrip memberi rasa kontrol.	menganggapnya satu-satunya cara mengetahui kualitas.
Eksperimen ML	Model memprioritaskan tes berisiko; sistem menyarankan perbaikan locator.	Kecemasan akan hilangnya keahlian dan kesulitan memeriksa rekomendasi.	Keberanian belajar; kerendahan hati untuk menguji model; komunikasi terbuka tentang keterbatasan.
Integrasi	Dashboard menampilkan alasan prioritas, confidence, dan bukti perubahan.	Peran penguji beralih dari eksekutor skrip ke evaluator risiko.	Akuntabilitas: keputusan rilis tetap memiliki pemilik manusia dan jejak audit.
Governance	Model dipantau untuk drift, kesalahan, keamanan, dan privasi data.	Tim lintas fungsi terlibat: QA, dev, data, security, product.	Keadilan, kepedulian pada pengguna, dan kolaborasi manusia-AI.

Kasus konseptual tersebut menggambarkan bahwa keberhasilan *AI-assisted SQA* bergantung pada tiga kondisi. Pertama, model harus dapat diuji secara teknis melalui metrik yang relevan, pengujian ketahanan, dan pemantauan *pascadeployment*. Kedua, rekomendasi perlu dapat dijelaskan pada tingkat yang memadai bagi pengambil keputusan. Ketiga, organisasi harus menata ulang peran agar AI memperluas kapasitas manusia, bukan menghapus tanggung jawab manusia. Prinsip ini selaras dengan NIST AI RMF yang menempatkan *governance* sebagai fungsi lintas-siklus dan memandang transparansi serta akuntabilitas sebagai elemen yang menembus seluruh karakteristik AI yang terpercaya (Tabassi, 2023).

### 3.4 *Technical Reluctance* sebagai Isu Pengembangan Karakter, Bukan Hambatan Individual

Pembacaan hermeneutika memperlihatkan bahwa *technical reluctance* perlu dipahami secara lebih manusiawi. Resistensi sering muncul ketika seseorang merasa kehilangan tiga hal: kompetensi yang telah dibangun lama, kepastian mengenai cara kerja yang benar, dan pengakuan atas identitas profesionalnya. Menyebut penguji sebagai ‘resisten terhadap perubahan’ tanpa menelaah pengalaman tersebut justru mengabaikan dimensi karakter yang hendak dibina. Karakter tidak berkembang melalui pemaksaan untuk menerima teknologi; ia berkembang melalui pengalaman bertanggung jawab dalam menghadapi ketidakpastian.

Maslow membantu mengubah fokus pendidikan organisasi. Program *reskilling* tidak seharusnya hanya mengajarkan penggunaan platform AI. Program tersebut perlu mengajak rekayasawan membedakan kapan ketergantungan pada skrip memberi keamanan yang sehat dan kapan ia berubah menjadi penghindaran terhadap masalah baru. Rekayasawan perlu diberi kesempatan untuk menguji model, mencari kegagalan model, membandingkan rekomendasi AI dengan penalaran domain, serta mendiskusikan *trade-off* secara terbuka. Praktik tersebut menumbuhkan *growth orientation* yang tidak naif: tumbuh bukan berarti percaya penuh pada AI, melainkan berani menilai AI secara kritis.

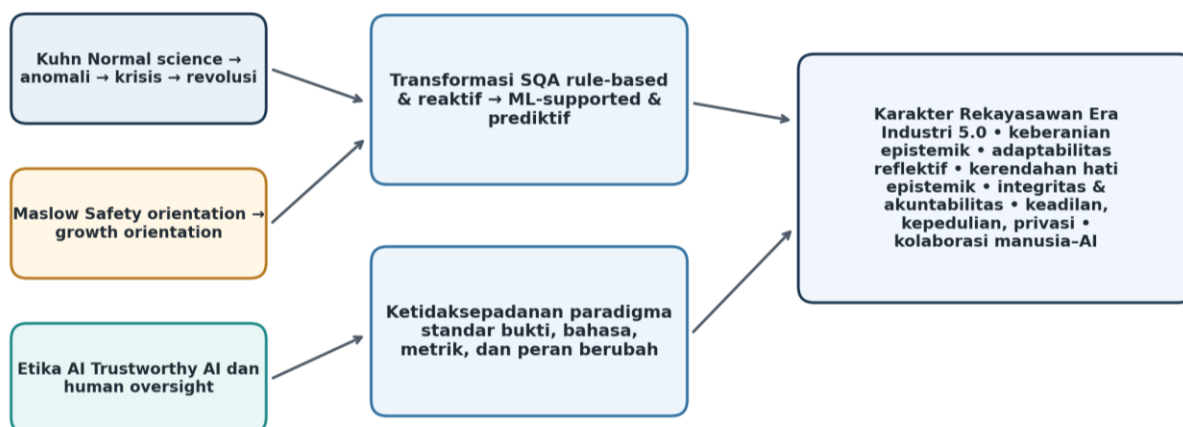
### 3.5 Model Karakter Rekayasawan Era Industri 5.0

Sintesis penelitian menghasilkan enam dimensi karakter yang diperlukan dalam *ML-supported SQA*. Model ini tidak dimaksudkan sebagai daftar sifat ideal yang abstrak. Setiap dimensi dikaitkan dengan tindakan yang dapat diamati dalam proses rekayasa. Dengan demikian, organisasi dapat menggunakannya untuk desain kurikulum, pengembangan kompetensi, evaluasi tata kelola, dan pembentukan budaya kerja.

**Tabel 4. Model Karakter Rekayasawan Era Industri 5.0**

Dimensi Karakter	Makna dalam SQA Berbasis ML	Contoh Perilaku Profesional
Keberanian epistemik	Berani meninjau ulang asumsi lama ketika bukti menunjukkan keterbatasan paradigma.	Mengusulkan eksperimen model, mengakui keterbatasan test suite, dan tidak menyembunyikan ketidakpastian.
Adaptabilitas reflektif	Mau belajar ulang tanpa menjadikan kebaruan sebagai tujuan itu sendiri.	Mengevaluasi apakah ML benar-benar diperlukan, sesuai risiko, dan memberi nilai tambah.
Kerendahan hati epistemik	Mengakui bahwa model, data, dan penilaian manusia dapat keliru.	Memeriksa bias, drift, dan kasus tepi; meminta second review pada keputusan berisiko.
Integritas dan akuntabilitas	Menjaga jejak keputusan dan tidak melepaskan tanggung jawab kepada sistem.	Mendokumentasikan data, versi model, alasan rilis, serta mekanisme rollback.
Keadilan, kepedulian, dan privasi	Memperhatikan dampak keputusan kualitas bagi pengguna yang berbeda.	Meminimalkan data sensitif, mengevaluasi bias, dan mengutamakan keselamatan pengguna.
Kolaborasi manusia-AI	Memosisikan AI sebagai mitra analitis, bukan otoritas yang tak dapat digugat.	Mendesain human-in-the-loop, forum lintas fungsi, dan mekanisme eskalasi keputusan.

Model konseptual berikut menyatukan dinamika Kuhn, Maslow, dan *Trustworthy AI*. Arah model menunjukkan bahwa revolusi teknis akan menghasilkan transformasi karakter hanya apabila organisasi memberikan pengawasan, pembelajaran, dan tata kelola yang memadai.



Gambar 1. Model konseptual transformasi karakter rekayasawan dalam *revolusi ML-supported SQA*

### 3.6 Trustworthy dan Explainable AI: Dari Prinsip ke Rutinitas Kerja

*Trustworthy AI* hanya bermakna apabila diterjemahkan menjadi rutinitas kerja. Dalam SQA, validitas dan reliabilitas berarti model diuji pada data yang relevan, dipantau ketika konteks berubah, dan dievaluasi terhadap kerugian yang mungkin timbul dari kesalahan prediksi. Transparansi berarti tim mengetahui tujuan model, sumber data, batas penggunaan, dan pihak yang bertanggung jawab. *Explainability* berarti terdapat alasan yang cukup untuk memahami faktor yang mendorong prioritas tes atau prediksi risiko. *Fairness* dan privasi berarti data serta dampak keputusan diperiksa agar tidak merugikan kelompok tertentu atau mengekspos informasi sensitif.

Tantangan etis yang penting adalah *trade-off*. Model yang lebih kompleks mungkin meningkatkan prediksi tetapi mengurangi keterjelasan alasan; teknik privasi mungkin mengurangi informasi yang tersedia untuk evaluasi; dan percepatan rilis dapat berbenturan dengan kebutuhan validasi yang memadai. NIST menekankan bahwa karakteristik trustworthiness harus diseimbangkan secara kontekstual, bukan diperlakukan sebagai kotak centang yang berdiri sendiri (Tabassi, 2023). Karena itu, karakter rekayasawan diperlukan untuk mengambil keputusan ketika tidak semua nilai dapat dimaksimalkan sekaligus.

Prinsip tersebut juga selaras dengan UNESCO yang menempatkan transparansi, *fairness*, dan *human oversight* sebagai syarat penghormatan terhadap martabat serta hak manusia (UNESCO, 2021). Pada level organisasi, *human oversight* perlu dipahami lebih luas daripada seorang pengawas yang menekan tombol persetujuan. Ia mencakup pembagian wewenang yang jelas, dokumentasi, mekanisme sanggahan, partisipasi pemangku kepentingan, dan budaya yang mengizinkan orang menghentikan proses ketika bukti keselamatan atau keadilan belum memadai.

### 3.7 Implikasi bagi Pendidikan Karakter, Organisasi, dan Penelitian

Bagi pendidikan tinggi, pembelajaran AI dan *software engineering* perlu mengintegrasikan latihan teknis dengan latihan penilaian etis. Mahasiswa tidak hanya diminta membangun model yang akurat, tetapi juga menjelaskan keterbatasannya, membuat *model card* sederhana, mengidentifikasi pihak yang berisiko terdampak, dan mempertahankan keputusan desain di hadapan skenario konflik nilai. Pendekatan ini memperlakukan karakter sebagai kebiasaan menalar dan bertindak, bukan sebagai slogan tambahan pada akhir mata kuliah.

Bagi organisasi, indikator keberhasilan transformasi digital perlu diperluas. Selain metrik *delivery* dan *defect rate*, organisasi dapat memantau keberadaan audit trail, proses review lintas fungsi, kualitas dokumentasi model, kemampuan anggota tim menjelaskan risiko, serta efektivitas forum eskalasi. Budaya yang sehat tidak menghukum orang karena mengungkapkan ketidakpastian; budaya tersebut menjadikan ketidakpastian sebagai bahan investigasi bersama.

Bagi penelitian lanjutan, model ini dapat diuji secara empiris melalui studi kasus multi-organisasi, wawancara dengan *QA engineer*, survei orientasi *safety-growth*, atau *action research* pada program *reskilling AI*. Penelitian juga dapat mengembangkan instrumen untuk mengukur enam dimensi karakter rekayasawan serta menguji hubungannya dengan kualitas keputusan, kepercayaan pada AI, dan perilaku etis dalam tim.

#### IV. KESIMPULAN

Artikel ini menegaskan bahwa revolusi *machine learning* dalam penjaminan kualitas perangkat lunak bukan hanya perubahan teknologi. Dengan lensa Kuhn, pergeseran dari pengujian *rule-based* yang deterministik menuju SQA berbasis pembelajaran adalah perubahan paradigma karena ia mengganti bentuk masalah, jenis bukti, ukuran keberhasilan, bahasa kerja, dan peran profesional. Konsep *incommensurability* menjelaskan mengapa perubahan tersebut tidak dapat diselesaikan dengan menambahkan alat baru pada proses lama.

Sintesis dengan Maslow memperlihatkan bahwa krisis paradigma mempunyai dimensi psikologis. Rekayasawan perlu bergerak dari orientasi keamanan yang semata mencari kepastian menuju orientasi pertumbuhan yang mampu belajar dari ambiguitas, tanpa kehilangan disiplin dan tanggung jawab. Perubahan ini bukan ajakan untuk menyerahkan penilaian manusia kepada algoritme, melainkan ajakan untuk menumbuhkan keberanian epistemik, adaptabilitas reflektif, kerendahan hati, integritas, keadilan, kepedulian, dan kemampuan berkolaborasi dengan AI.

Kebaruan utama penelitian ini adalah perumusan Kerangka Karakter Rekayasawan Era Industri 5.0 melalui sintesis filsafat sains Kuhn, psikologi humanistik Maslow, dan tata kelola *Trustworthy AI*. Kerangka tersebut menempatkan *Explainable AI* dan *human oversight* sebagai fondasi moral-operasional, bukan aksesori kepatuhan. Ketika AI ikut memengaruhi keputusan kualitas, manusia tetap memikul tanggung jawab untuk memahami konteks, mengevaluasi risiko, melindungi pengguna, dan mempertanggungjawabkan keputusan. Dengan demikian, masa depan SQA yang matang bukanlah masa ketika manusia tersingkir dari kualitas, melainkan masa ketika rekayasawan bertumbuh menjadi penjaga kualitas yang lebih reflektif, adil, dan bertanggung jawab.

#### REFERENSI

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>

- Avşar, Y. (2025). The evolution of quality management in the era of Industry 4.0 and 5.0. *Archives of Engineering Knowledge*, 11(1), 77-83. <https://doi.org/10.30657/aek.2025.11.14>
- Batool, I., & Khan, T. A. (2023). Software fault prediction using deep learning techniques. *Software Quality Journal*, 31(4), 1241-1280. <https://doi.org/10.1007/s11219-023-09642-4>
- Floridi, L., Cowsls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Gadamer, H.-G. (2004). *Truth and method* (2nd rev. ed.; J. Weinsheimer & D. G. Marshall, Trans.). Continuum.
- Gomaa, A. H. (2025). Quality management excellence in the era of Industry 4.0 (Quality 4.0): A comprehensive review, gap analysis, and strategic framework. *Advancements in Science and Technology*, 2, 1-40.
- Hagendorff, T. (2022). A virtue-based framework to support putting AI ethics into practice. *Philosophy & Technology*, 35, Article 55. <https://doi.org/10.1007/s13347-022-00553-z>
- Hagendorff, T. (2023). AI ethics and its pitfalls: Not living up to its own standards? *AI and Ethics*, 3, 329-336. <https://doi.org/10.1007/s43681-022-00173-5>
- Huang, J. (2023). Digital engineering transformation with trustworthy AI towards Industry 4.0: Emerging paradigm shifts. *Journal of Integrated Design and Process Science*, 26(3-4), 267-290. <https://doi.org/10.3233/JID-229010>
- Jabareen, Y. (2009). Building a conceptual framework: Philosophy, definitions, and procedure. *International Journal of Qualitative Methods*, 8(4), 49-62. <https://doi.org/10.1177/160940690900800406>
- Khanam, R., Hussain, M., Hill, R., & Allen, P. (2024). A comprehensive review of convolutional neural networks for defect detection in industrial applications. *IEEE Access*, 12, 94250-94295. <https://doi.org/10.1109/ACCESS.2024.3425166>
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—A systematic literature review. *Information and Software Technology*, 51(1), 7-15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- Kožnjak, B. (2017). Kuhn meets Maslow: The psychology behind scientific revolutions. *Journal for General Philosophy of Science*, 48(2), 257-287. <https://doi.org/10.1007/s10838-016-9352-x>
- Kuhn, T. S. (2012). *The structure of scientific revolutions: 50th anniversary edition*. University of Chicago Press.
- Maslow, A. H. (1966). *The psychology of science: A reconnaissance*. Harper & Row.
- Ngo, K., Nguyen, V., & Nguyen, T. (2022). Research on test flakiness: From unit to system testing. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*. <https://doi.org/10.1145/3551349.3563242>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pan, R., Bagherzadeh, M., Ghaleb, T. A., & Briand, L. (2022). Test case selection and prioritization using machine learning: A systematic literature review. *Empirical Software Engineering*, 27, Article 29. <https://doi.org/10.1007/s10664-021-10066-6>

- Pham, P., Nguyen, V., & Nguyen, T. (2022). A review of AI-augmented end-to-end test automation tools. In Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering (pp. 1-4). <https://doi.org/10.1145/3551349.3563240>
- Ricca, F., Marchetto, A., & Stocco, A. (2025). A multi-year grey literature review on AI-assisted test automation. *Information and Software Technology*, 186, 107799. <https://doi.org/10.1016/j.infsof.2025.107799>
- Sønning, L., & Werner, V. (2021). The replication crisis, scientific revolutions, and linguistics. *Linguistics*, 59(5), 1179-1206. <https://doi.org/10.1515/ling-2019-0045>
- Tabassi, E. (Ed.). (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- UNESCO. (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization.
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
- Wang, S., Huang, L., Gao, A., Ge, J., Zhang, T., Feng, H., Satyarth, I., Li, M., Zhang, H., & Ng, V. (2023). Machine/deep learning for software engineering: A systematic literature review. *IEEE Transactions on Software Engineering*, 49(3), 1188-1231. <https://doi.org/10.1109/TSE.2022.3173346>
- Yao, X., Ma, N., Zhang, J., Wang, K., Yang, E., & Faccio, M. (2024). Enhancing wisdom manufacturing as industrial metaverse for Industry and Society 5.0. *Journal of Intelligent Manufacturing*, 35(1), 235-255. <https://doi.org/10.1007/s10845-022-02027-7>