PANDUAN MANAJEMEN RISIKO **AI**

UNTUK PELAKSANA MANAJEMEN RISIKO AI

LATAR BELAKANG

Perkembangan teknologi kecerdasan buatan (Artificial Intelligence / AI) telah membawa transformasi besar dalam berbagai sektor industri, mulai dari keuangan, kesehatan, manufaktur, hingga layanan publik. Penerapan AI memungkinkan efisiensi proses, peningkatan akurasi pengambilan keputusan, dan penciptaan nilai bisnis baru. Namun, di balik potensi manfaat tersebut, penggunaan AI juga membawa berbagai risiko yang bersifat teknis, etis, sosial, hukum, dan operasional.

Berbeda dengan sistem konvensional, AI bersifat dinamis, otonom, dan berbasis data besar yang kompleks. Hal ini menjadikan AI rentan terhadap risiko seperti bias algoritma, kurangnya transparansi (black box), pelanggaran privasi, dan potensi diskriminasi terhadap kelompok tertentu. Selain itu, penggunaan AI yang tidak tepat dapat menimbulkan dampak reputasional, hukum, hingga hilangnya kepercayaan dari pemangku kepentingan.

Menyadari hal ini, berbagai lembaga internasional telah mengembangkan kerangka kerja dan standar untuk mengelola risiko AI secara sistematis dan bertanggung jawab, seperti ISO/IEC 23894:2023, NIST AI Risk Management Framework (AI RMF 1.0), serta prinsip-prinsip dari OECD dan regulasi Uni Eropa (EU AI Act).

Panduan ini bertujuan untuk membantu organisasi dalam mengidentifikasi, menganalisis, mengevaluasi, mengendalikan, dan memantau risiko terkait penggunaan AI secara efektif dan berkelanjutan, sehingga dapat memastikan bahwa sistem AI yang digunakan aman, andal, etis, dan mematuhi regulasi yang berlaku.

DASAR PEMIKIRAN

Penggunaan kecerdasan buatan *(Artificial Intelligence / AI)* di lingkungan perusahaan tidak lagi bersifat opsional, melainkan menjadi elemen strategis yang mendorong inovasi dan keunggulan kompetitif. Namun, sifat AI yang kompleks, tidak sepenuhnya dapat diprediksi, dan terus belajar dari data membuat teknologi ini memiliki karakter risiko yang berbeda dibandingkan sistem informasi konvensional.

Tanpa pendekatan manajemen risiko yang terstruktur, penerapan AI berisiko menghasilkan keputusan yang tidak adil, menimbulkan dampak hukum, merusak reputasi organisasi, atau

bahkan membahayakan keselamatan pengguna. Oleh karena itu, dibutuhkan panduan internal yang dapat mengarahkan perusahaan dalam menerapkan prinsip kehati-hatian, akuntabilitas, dan keandalan dalam pengembangan maupun penggunaan sistem AI.

Dokumen ini disusun sebagai respons terhadap kebutuhan tersebut, dengan merujuk pada kerangka kerja dan standar global yang kredibel. Panduan ini bertujuan untuk:

- Memberikan struktur dan langkah praktis dalam mengelola risiko AI,
- Meningkatkan kesadaran dan kapasitas internal dalam mengidentifikasi serta menangani potensi risiko,
- Memastikan bahwa penggunaan AI sesuai dengan nilai etika dan prinsip tata kelola yang baik,
- Mempersiapkan organisasi untuk mematuhi regulasi yang terus berkembang, seperti UU PDP dan UU ITE.

Dengan memiliki panduan ini, perusahaan dapat lebih percaya diri dalam memanfaatkan AI secara bertanggung jawab dan berkelanjutan, sembari melindungi kepentingan pengguna, pemangku kepentingan, dan reputasi organisasi.

TUJUAN DAN RUANG LINGKUP

Panduan ini bertujuan memberikan kerangka sistematis bagi perusahaan dalam mengidentifikasi, menilai, mengelola, dan memantau risiko terkait penggunaan sistem AI, guna memastikan bahwa AI digunakan secara aman, bertanggung jawab, dan sesuai regulasi.

PRINSIP DASAR

Prinsip-prinsip dasar manajemen risiko AI disusun dengan mengacu pada berbagai standar dan kerangka kerja kredibel seperti ISO/IEC 23894:2023, NIST AI Risk Management Framework, serta OECD AI Principles. Penjabarannya prinsip-prinsip dasar manajemen risiko AI adalah berikut:

1. **Transparansi**

   Sistem AI harus dapat dijelaskan secara proporsional kepada pemangku kepentingan. Ini mencakup keterbukaan terhadap tujuan sistem, logika dasar model, serta data pelatihan yang digunakan. Transparansi membantu mengurangi ketidakpastian dan memperkuat kepercayaan pengguna.

2. **Keamanan dan Keandalan *(Safety & Robustness)***

   Sistem AI harus dirancang untuk tahan terhadap kesalahan, gangguan, dan penyalahgunaan. Hal ini mencakup evaluasi keamanan siber, pengujian performa terhadap skenario ekstrem, serta mekanisme deteksi kesalahan atau *"fail-safe"*.

3. **Keadilan dan Non-diskriminasi *(Fairness & Non-discrimination)***

   Sistem tidak boleh mendiskriminasi berdasarkan ras, gender, usia, atau atribut sensitif lainnya. Risiko bias harus diidentifikasi dan diminimalkan sejak tahap desain dan pelatihan model.

4. **Akuntabilitas**

   Tanggung jawab atas keluaran dan dampak sistem AI harus dapat ditelusuri ke pengembang, penyedia, atau pengguna sistem. Audit trail, dokumentasi, dan pengawasan internal menjadi kunci untuk akuntabilitas.

5. **Kepatuhan terhadap Hukum dan Etika**

   AI harus mematuhi peraturan yang berlaku seperti **UU Pelindungan Data Pribadi (UU PDP)**, serta ketentuan lokal dan internasional lainnya. Selain aspek hukum, prinsip etika perlu dijadikan landasan dalam seluruh siklus hidup AI.

6. **Keselarasan dengan Nilai Organisasi**

   Risiko AI harus dikelola sejalan dengan visi, misi, dan nilai organisasi. Ini memastikan bahwa teknologi tidak hanya berfungsi secara efisien, tetapi juga mendukung tujuan jangka panjang perusahaan.

Dengan prinsip-prinsip tersebut sebagai fondasi, panduan ini bertujuan menyediakan pendekatan praktis yang dapat diintegrasikan dalam pengembangan, implementasi, dan pengawasan sistem AI di berbagai unit perusahaan. Panduan ini juga menjadi alat bantu untuk meningkatkan kesiapan organisasi dalam menghadapi audit, pengawasan, atau persyaratan kepatuhan di masa mendatang.

KERANGKA RISIKO AI

Manajemen risiko AI merupakan proses sistematis yang dirancang untuk membantu organisasi dalam mengidentifikasi, menganalisis, mengevaluasi, mengendalikan, dan memantau risiko yang timbul dari penggunaan sistem berbasis kecerdasan buatan. Kerangka ini mengikuti prinsip-prinsip manajemen risiko umum seperti yang diatur dalam ISO 31000, dan disesuaikan dengan karakteristik unik teknologi AI sebagaimana tertuang dalam ISO/IEC 23894 dan NIST AI RMF 1.0.

KONTEKS DAN TATA KELOLA

Langkah awal adalah menetapkan konteks internal dan eksternal yang memengaruhi penggunaan AI dalam organisasi (visi dan misi). Ini mencakup pemetaan pemangku kepentingan, peraturan yang relevan, serta struktur tata kelola risiko dan etika AI seperti Komite Risiko AI, Etika & Kepatuhan, Penanggung jawab risiko AI.

IDENTIFIKASI RISIKO

Organisasi perlu mengidentifikasi berbagai jenis risiko AI yang mungkin timbul, mulai dari risiko teknis (seperti bias algoritma atau *adversarial attack*), risiko etis dan sosial (seperti diskriminasi atau pelanggaran privasi), risiko hukum (ketidakpatuhan terhadap UU PDP), operasioal (ketergantungan sistem, error pada *deployment*), hingga reputasi (kepercayaan publik menurun).

ANALISIS DAN EVALUASI RISIKO

Risiko yang telah diidentifikasi dianalisis untuk menilai kemungkinan terjadinya dan dampaknya terhadap organisasi. Risiko kemudian dievaluasi menggunakan pendekatan kuantitatif (skala 1–5) atau kualitatif (rendah-sedang-tinggi), sehingga dapat diprioritaskan secara proporsional.

PENGENDALIAN RISIKO

Organisasi harus menerapkan langkah-langkah mitigasi yang tepat yaitu antara lain:

- Desain aman dan andal *(secure & robust by design)*
- Pengujian dan validasi model
- Audit algoritma dan data
- Explainability & dokumentasi
- Red teaming dan adversarial testing

- Model card dan data card
- Implementasi fallback atau *human-in-the-loop* (HITL)

PEMANTAUAN DAN REVIEW

Risiko AI bersifat dinamis karena data dan model AI terus berubah. Oleh karena itu, perlu ada proses pemantauan performa model, deteksi perubahan konteks (seperti data drift), serta tinjauan berkala terhadap efektivitas mitigasi yang diterapkan.

Kerangka ini dirancang untuk bersifat fleksibel dan adaptif terhadap kebutuhan organisasi, serta dapat diintegrasikan ke dalam proses tata kelola teknologi dan manajemen risiko korporat yang telah ada. Pendekatan ini mendukung penerapan AI secara bertanggung jawab, berkelanjutan, dan sesuai regulasi.

KLASIFIKASI RISIKO AI

Sebagai pelengkap pendekatan berbasis kerangka manajemen risiko, klasifikasi risiko AI dapat digunakan untuk membantu perusahaan menilai tingkat pengawasan dan pengendalian yang sesuai terhadap sistem AI yang digunakan. Klasifikasi ini membagi sistem AI ke dalam empat kategori berdasarkan tingkat risikonya terhadap hak asasi manusia, keselamatan, dan nilai-nilai sosial.

UNACCEPTABLE RISK (RISIKO TAK DAPAT DITERIMA)

Kategori ini mencakup sistem AI yang dianggap bertentangan dengan nilai-nilai fundamental dan secara inheren berbahaya terhadap hak individu. Sistem dalam kategori ini dilarang secara langsung oleh hukum Uni Eropa.

Contoh:

- Sistem AI untuk *social scoring* oleh pemerintah seperti yang digunakan dalam sistem pemantauan sosial massal.
- Sistem yang mengeksploitasi kerentanan anak-anak atau individu dengan disabilitas mental.
- AI yang memanipulasi perilaku melalui teknik subliminal secara tidak etis.

Tindakan: Sistem dalam kategori ini tidak boleh digunakan dan harus dieliminasi dari pipeline pengembangan.

HIGH RISK (RISIKO TINGGI)

Sistem AI diklasifikasikan sebagai risiko tinggi jika digunakan dalam domain sensitif yang dapat mempengaruhi keselamatan, hak fundamental, atau kehidupan seseorang secara langsung. Sistem ini diperbolehkan, tetapi harus mematuhi persyaratan regulatif yang sangat ketat, termasuk penilaian risiko, dokumentasi teknis, audit, dan pengawasan manusia.

Contoh:

- Sistem rekrutmen yang secara otomatis menyaring atau memberi skor kandidat.
- Sistem evaluasi kinerja karyawan.
- AI dalam sistem pendidikan yang memengaruhi akses terhadap pendidikan.
- Sistem AI dalam penegakan hukum atau sistem identifikasi biometrik di ruang publik.

Tindakan: Harus menjalani uji ketat, registrasi, audit, dan evaluasi sebelum diterapkan.

LIMITED RISK (RISIKO TERBATAS)

Sistem AI dengan risiko terbatas adalah sistem yang interaksinya dengan manusia bersifat langsung dan relatif tidak berdampak signifikan terhadap keselamatan atau hak individu, namun tetap memerlukan transparansi.

Contoh:

- Chatbot layanan pelanggan.
- Sistem AI yang merekomendasikan konten pada *platform e-commerce* atau media sosial.

Tindakan: Harus menyatakan bahwa pengguna sedang berinteraksi dengan AI, misalnya melalui pemberitahuan eksplisit atau label.

MINIMAL RISK (RISIKO MINIMAL)

Sistem AI dalam kategori ini menimbulkan risiko yang sangat rendah atau nyaris tidak ada. Sistem seperti ini dapat digunakan secara bebas dan tanpa persyaratan tambahan.

Contoh:

- AI untuk filter spam email.
- Sistem rekomendasi film atau musik yang bersifat personalisasi.

Tindakan: Tidak memerlukan pengawasan tambahan, cukup mengikuti prinsip umum tata kelola AI yang baik.

Klasifikasi ini dapat digunakan sebagai alat bantu awal dalam proses identifikasi dan prioritisasi risiko AI, terutama saat melakukan pemetaan use case AI dalam organisasi. Dengan menetapkan kategori risiko sejak awal, perusahaan dapat menyesuaikan tingkat mitigasi, dokumentasi, dan pengawasan yang sesuai sebelum sistem digunakan secara operasional.

DOKUMENTASI

Salah satu pilar utama dalam manajemen risiko AI yang efektif adalah dokumentasi yang komprehensif dan jejak audit (*audit trail*) yang dapat ditelusuri. Dokumentasi dan audit trail berfungsi sebagai fondasi untuk transparansi, akuntabilitas, dan kepatuhan, baik secara internal (tata kelola perusahaan) maupun eksternal (regulasi).

TUJUAN DOKUMENTASI DAN AUDIT TRAIL

- Memastikan bahwa setiap keputusan teknis, etis, dan bisnis terkait AI dapat dipertanggungjawabkan.
- Memudahkan proses audit, evaluasi performa, dan investigasi insiden jika terjadi masalah.
- Memenuhi persyaratan regulasi seperti UU Pelindungan Data Pribadi (UU PDP), dan potensi audit dari pihak ketiga.
- Mendukung budaya organisasi yang berbasis transparansi dan tata kelola yang baik (*good governance*).

ELEMEN DOKUMENTASI YANG DISARANKAN

Berikut adalah komponen dokumentasi minimum yang direkomendasikan untuk setiap sistem AI yang dikembangkan atau diadopsi:

1. ***Model Card*.** Dokumen yang menjelaskan rincian teknis dari model AI, termasuk:
   - Arsitektur model
   - Dataset pelatihan dan validasi
   - Tujuan penggunaan *(intended use)*
   - Keterbatasan *(limitations)*
   - Evaluasi performa dan fairness
2. ***Data Card*.** Dokumentasi yang menjelaskan karakteristik dataset yang digunakan:
   - Sumber data
   - Proses pembersihan dan transformasi

- Representasi demografis
- Potensi bias dan upaya mitigasinya

3. **Dokumen Evaluasi Risiko**
   - Identifikasi risiko utama berdasarkan *use case*
   - Penilaian kemungkinan dan dampak
   - Tindakan mitigasi yang diterapkan
   - Justifikasi keputusan
4. ***Catatan Validasi dan Uji Coba***
   - Hasil uji fungsional dan pengujian teknis
   - *Red teaming* atau *adversarial testing*
   - Review dari tim independen (jika ada)
5. **Log Aktivitas dan Perubahan Sistem**
   - Riwayat pelatihan model
   - Riwayat deployment dan update sistem
   - Perubahan parameter dan data training
   - Catatan siapa melakukan apa dan kapan *(audit trail)*
6. **Dokumentasi Tata Kelola**
   - Siapa yang bertanggung jawab atas sistem *(role & accountability)*
   - Struktur pengambilan keputusan
   - Proses eskalasi dan mitigasi jika terjadi insiden

AUDIT TRAIL

Audit trail berfungsi sebagai "rekam jejak" digital atas seluruh aktivitas terkait sistem AI. Ini sangat penting untuk:

- Penelusuran keputusan otomatis, terutama jika output AI berdampak langsung pada individu (misalnya dalam rekrutmen, pemberian kredit, atau penilaian kinerja).
- Membuktikan kepatuhan hukum, terutama jika sistem menghadapi tuntutan hukum atau pengawasan regulator.

- Evaluasi dan *continuous improvement,* karena data historis dapat digunakan untuk perbaikan model dan mitigasi berulang.

BEST PRACTICE DALAM DOKUMENTASI & AUDIT

- Gunakan template standar (misalnya *model cards* dan *data cards* dari Google/Meta) untuk mempermudah replikasi.
- Simpan dokumentasi secara digital dalam sistem terpusat dan versi terkontrol.
- Buat mekanisme review berkala atas dokumentasi dan *audit trail* oleh tim *independen*.
- Pastikan dokumentasi dapat diakses oleh tim legal, teknis, dan manajerial dengan kontrol hak akses yang jelas.

Dokumentasi dan *audit trail* bukan sekadar formalitas administratif, melainkan bagian krusial dari trust architecture dalam penggunaan AI. Dokumentasi yang baik akan memperkuat kepercayaan pemangku kepentingan terhadap sistem AI, serta meminimalkan risiko di masa depan.

PERAN & TANGGUNG JAWAB

Keberhasilan manajemen risiko AI tidak hanya bergantung pada teknologi dan prosedur, tetapi juga pada keterlibatan aktor-aktor kunci di seluruh organisasi. Setiap pihak memiliki peran dan tanggung jawab yang berbeda namun saling melengkapi, baik dalam tahap desain, implementasi, maupun evaluasi sistem AI. Pendekatan ini sejalan dengan prinsip governance by design, yang menekankan pentingnya kolaborasi lintas fungsi dalam pengelolaan teknologi kritikal.

1. **Tim Data Science / AI Engineer**

   Tanggung jawab utama:

   - Mendesain dan membangun model AI yang aman, transparan, dan adil.
   - Melakukan validasi, dokumentasi teknis *(model card),* dan analisis bias.
   - Menerapkan prinsip *secure & robust by design* dalam pengembangan model.
   - Berkoordinasi dengan tim lain untuk memahami dampak sosial dan regulatif.

2. **Tim Legal / Kepatuhan (*Compliance*)**

   Tanggung jawab utama:

   - Memastikan sistem AI mematuhi regulasi yang berlaku (misalnya UU PDP).
   - Mengkaji kontrak dan perjanjian vendor terkait penggunaan data dan AI.
   - Menilai risiko hukum dari penggunaan sistem yang menghasilkan keputusan otomatis.
   - Memonitor perkembangan hukum terkait AI secara berkala.

3. **Komite Etika AI / *Governance Board***

   Tanggung jawab utama:

   - Meninjau risiko sosial dan etis dari sistem AI, termasuk fairness, diskriminasi, dan privasi.
   - Menetapkan kebijakan internal terkait pengembangan dan penggunaan AI.
   - Menjadi forum eskalasi jika muncul isu yang bersifat multidisipliner.
   - Memberikan panduan strategis terkait adopsi AI yang bertanggung jawab.

4. ***Chief Information Security Officer (CISO)* / Tim Keamanan**

   Tanggung jawab utama:

   - Menilai keamanan sistem dan potensi serangan terhadap model (misalnya *adversarial attack)*.
   - Mengembangkan kontrol akses dan keamanan data yang digunakan dalam pelatihan AI.
   - Memastikan AI tidak menjadi vektor baru untuk kerentanan siber.
   - Mengintegrasikan kontrol AI ke dalam kebijakan keamanan TI perusahaan.

5. **Tim *Human Resources* / SDM**

   Tanggung jawab utama (khusus jika AI digunakan dalam konteks SDM):

   - Memastikan sistem AI untuk rekrutmen, penilaian kinerja, atau promosi dilakukan secara adil dan inklusif.
   - Memberikan pelatihan kepada karyawan tentang penggunaan dan dampak AI.
   - Mendorong budaya keterbukaan terhadap AI dan pelaporan insiden.

6. **Manajemen Senior / Direksi**

   Tanggung jawab utama:

   - Menetapkan arah strategis dan toleransi risiko terhadap teknologi AI.
   - Mengalokasikan sumber daya untuk pengembangan, pemantauan, dan mitigasi risiko.
   - Bertanggung jawab penuh atas dampak eksternal dan reputasi dari adopsi AI.
   - Mendorong tata kelola yang terintegrasi antara AI, risiko, dan keberlanjutan.

PRINSIP UMUM DISTRIBUSI TANGGUNG JAWAB:

- Tanggung jawab teknis tidak dapat dipisahkan dari tanggung jawab etis dan hukum.
- Keterlibatan multi-disiplin adalah kunci: AI bukan hanya domain teknolog.
- Harus ada pengambil keputusan yang jelas untuk setiap tahap siklus hidup AI.
- Dokumentasi tentang siapa melakukan apa dan kapan (RACI chart) sangat dianjurkan.

PENUTUP

Panduan ini bersifat dinamis dan perlu diperbarui sesuai dengan perkembangan teknologi AI dan regulasi global. Perusahaan didorong untuk melakukan benchmarking dan mengadopsi praktik terbaik dari industri.

LAMPIRAN

TEMPLATE PENILAIAN RISIKO AI

Berdasarkan kemungkinan, dampak, dan tingkat risiko secara keseluruhan. Gunakan skala 1–5 untuk penilaian kuantitatif, atau kategori (Rendah, Sedang, Tinggi) untuk penilaian kualitatif.

Use Case / Sistem AI	Jenis Risiko	Deskripsi Risiko	Probabilitas (1 – 5)	Dampak (1 – 5)	Tingkat Risiko

Catatan:

- **Tingkat Risiko = Probabilitas x Dampak**
- **Skor 1–5: 1=Rendah, 5=Sangat Tinggi**
- **Lakukan mitigasi terhadap risiko dengan skor tinggi (15 ke atas)**

Berikut adalah contoh pengisian template penilaian risiko AI untuk sistem rekrutmen berbasis AI yang digunakan untuk menyaring kandidat secara otomatis.

Use Case / Sistem AI	Jenis Risiko	Deskripsi Risiko	Probabilitas (1 – 5)	Dampak (1 – 5)	Tingkat Risiko
Sistem Penyaringan Kandidat AI	Diskriminasi	Model menunjukkan bias terhadap jenis kelamin tertentu dalam seleksi kandidat.	4	5	20

PEDOMAN SKOR RISIKO AI

Gunakan panduan berikut untuk menilai kemungkinan dan dampak risiko.

Skala Penilaian Kemungkinan dan Dampak:

Skor	Kategori	Deskripsi
1	Sangat rendah	Jarang terjadi / Dampak minimal
2	Rendah	Kemungkinan kecil / Dampak kecil
3	Sedang	Cukup mungkin terjadi / Dampak sedang
4	Tinggi	Kemungkinan besar terjadi / Dampak besar
5	Sangat Tinggi	Hampir pasti terjadi / Dampak kritikal

Interpretasi Tingkat Risiko (Skor Probabilitas x Dampak):

- 1–6 : Risiko Rendah – Dapat diterima tanpa tindakan signifikan.
- 8–12 : Risiko Sedang – Perlu mitigasi proporsional.
- 15–25: Risiko Tinggi – Butuh perhatian segera dan mitigasi kuat.