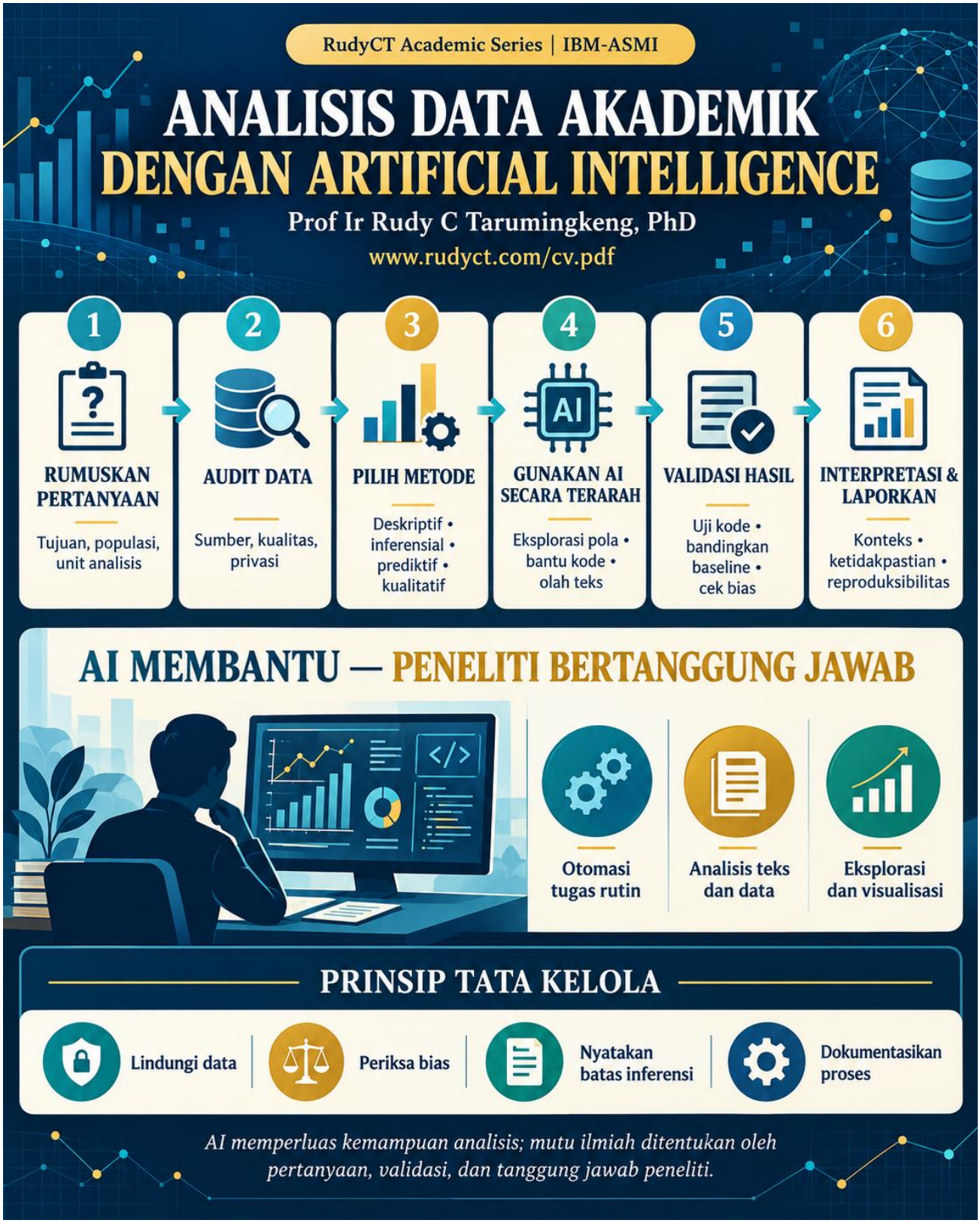

RudyCT Academic Series | IBM-ASMI
ANALISIS DATA AKADEMIK
DENGAN ARTIFICIAL INTELLIGENCE
Prof Ir Rudy C Tarumingkeng, PhD
www.rudyct.com/cv.pdf
1
RUMUSKAN PERTANYAAN
Tujuan, populasi, unit analisis
2
AUDIT DATA
Sumber, kualitas, privasi
3
PILIH METODE
Deskriptif • inferensial • prediktif • kualitatif
4
GUNAKAN AI SECARA TERARAH
Eksplorasi pola • bantu kode • olah teks
5
VALIDASI HASIL
Uji kode • bandingkan baseline • cek bias
6
INTERPRETASI & LAPORKAN
Konteks • ketidakpastian • reproduksibilitas
AI MEMBANTU — PENELITI BERTANGGUNG JAWAB
Otomasi tugas rutin
Analisis teks dan data
Eksplorasi dan visualisasi
PRINSIP TATA KELOLA
Lindungi data
Periksa bias
Nyatakan batas inferensi
Dokumentasikan proses
AI memperluas kemampuan analisis; mutu ilmiah ditentukan oleh pertanyaan, validasi, dan tanggung jawab peneliti.

Analisis Data Akademik dengan Artificial Intelligence

Oleh Prof Ir Rudy C Tarumingkeng, PhD

www.rudyct.com/cv.pdf

IBM-ASMI* | *RudyCT Academic Series* | *2026

Abstrak

Artikel ini menguraikan bagaimana artificial intelligence (AI), termasuk machine learning dan generative AI, dapat mendukung analisis data akademik tanpa mengurangi tanggung jawab ilmiah peneliti. Pembahasan mencakup perumusan tujuan analisis, pemahaman struktur data, pembersihan, eksplorasi, statistik inferensial, pemodelan prediktif, analisis teks kualitatif, integrasi bukti, visualisasi, interpretasi, dan pelaporan. AI dapat menurunkan hambatan teknis, membantu pemrograman serta pemrosesan data berskala besar, dan memperluas eksplorasi pola. Namun, otomasi dapat memperbesar kesalahan ketika definisi variabel, asumsi, kualitas sampel, atau konteks tidak dipahami. Model dapat menghasilkan kode yang keliru, kesimpulan yang melampaui desain, bias terhadap kelompok tertentu, dan prediksi yang tidak dapat direproduksi. Artikel mengusulkan kerangka analisis berpusat pada pertanyaan dan bukti: tetapkan estimand atau tujuan analisis, audit sumber dan kualitas data, pilih metode yang sesuai, validasi hasil dengan prosedur independen, kaji ketidakpastian dan keadilan, dokumentasikan pipeline, serta komunikasikan batas inferensi. Contoh penelitian pendidikan dan manajemen menunjukkan bahwa AI paling bermanfaat sebagai asisten yang transparan dan dapat diuji.

Keputusan statistik, interpretasi substantif, perlindungan data, dan pertanggungjawaban atas klaim tetap berada pada peneliti.

Kata Kunci

analisis data akademik; artificial intelligence; machine learning; generative AI; statistika; learning analytics; kualitas data; reproduksibilitas

1. Pendahuluan

Analisis data adalah kegiatan mengubah catatan, pengukuran, dokumen, interaksi, atau jejak digital menjadi bukti yang dapat menjawab pertanyaan penelitian. Ia bukan tahapan mekanis yang dimulai setelah data dikumpulkan. Sejak peneliti menentukan siapa yang diamati, konsep apa yang diukur, bagaimana kategori dibentuk, dan kasus mana yang dianggap layak, keputusan analitis telah ikut membentuk data. Artificial intelligence (AI) memperluas kemampuan untuk memproses dan menghubungkan informasi, tetapi juga menambahkan lapisan transformasi yang perlu dipahami. Pertanyaan sentral bukan hanya apakah sistem dapat menghitung, melainkan apakah analisis yang dihasilkannya sesuai dengan desain dan menghasilkan inferensi yang bisa dipertanggungjawabkan.

AI dalam analisis data akademik mencakup keluarga metode dan alat yang luas. Machine learning mempelajari pola untuk klasifikasi, regresi, pengelompokan, deteksi anomali, atau representasi. Natural language processing mengolah bahasa. Sistem generatif dapat menulis kode, merangkum data, memberi saran visualisasi, dan menyusun interpretasi awal. Perangkat analitik yang menyematkan AI mungkin mengotomatisasi pemilihan model atau menghasilkan narasi laporan. Semua itu tidak sama. Sebuah algoritma regresi teratur dengan parameter eksplisit berbeda dari chatbot yang menerima instruksi bahasa alami; model bahasa yang menulis kode

berbeda dari model prediksi yang dilatih pada data berlabel. Peneliti perlu mengetahui jenis fungsi yang sedang digunakan.

Manfaat AI terlihat pada penanganan dataset besar, pengolahan teks, eksplorasi interaksi kompleks, pembuatan kode, dan penyusunan ringkasan. Namun, kecepatan bukan jaminan validitas. Sebuah sistem dapat menghasilkan tabel rapi dari variabel yang salah dikodekan. Model prediktif dapat mencapai akurasi tinggi karena kebocoran data. Ringkasan bahasa dapat mengubah korelasi menjadi sebab-akibat. Kode yang tampak benar dapat memakai uji statistik yang tidak sesuai. Tantangan ini bukan alasan menolak seluruh teknologi; ia alasan untuk menjadikan validasi bagian dari desain analisis.

Artikel ini membahas penggunaan AI dalam analisis kuantitatif, kualitatif, metode campuran, dan data pendidikan serta manajemen. Kerangka yang diajukan bergerak dari pertanyaan dan tujuan inferensi menuju audit data, pemilihan metode, validasi, interpretasi, dan pelaporan. Setiap tahap mempunyai keputusan manusia yang tidak dapat digantikan oleh keluaran otomatis semata. Studi eksploratif tentang ChatGPT sebagai analis kuantitatif menunjukkan bahwa bantuan model dapat menurunkan ambang operasional, tetapi kinerja berbeda menurut tugas, perangkat lunak, instruksi, serta keahlian pengguna (Prandner et al., 2025). Bukti tersebut mendorong penggunaan yang terbuka dan terbatas sesuai kemampuan teruji.

2. Memahami istilah dan posisi AI dalam analisis

AI adalah istilah payung, bukan nama satu metode. Dalam praktik data akademik, istilah ini dapat merujuk kepada algoritma statistik, model machine learning, sistem rekomendasi, pemroses bahasa, model generatif, agen yang menggunakan perangkat, atau sistem pendukung keputusan. Perbedaan pentingnya adalah tugas dan mekanisme. Regresi linear mengestimasi hubungan di bawah struktur matematis yang dikenal; random forest

membentuk banyak pohon untuk memprediksi; model bahasa memprediksi token berikutnya dan dapat menyusun jawaban; sistem retrieval augmented generation mengambil dokumen dan merangkai respons berbasis dokumen tersebut. Peneliti perlu menyebutkan alat dan peran yang nyata, bukan menulis "AI menganalisis data" sebagai penjelasan penuh.

Dalam tulisan ini, "analisis dengan AI" mempunyai tiga bentuk. Pertama, AI sebagai objek penelitian, misalnya peneliti mengevaluasi akurasi model deteksi putus studi. Kedua, AI sebagai metode analisis, misalnya machine learning digunakan memprediksi risiko kelulusan. Ketiga, AI sebagai asisten alur kerja, misalnya chatbot membantu menulis sintaks R atau memberi ide visualisasi, sementara analisis final dijalankan dan diperiksa peneliti. Ketiga bentuk ini memerlukan standar pelaporan dan validasi berbeda. Jika AI hanya memperbaiki tata bahasa, pengaruhnya pada inferensi kecil; jika model mengklasifikasikan responden dan menentukan kesimpulan, pengaruhnya substantif.

Generative AI tidak identik dengan machine learning secara keseluruhan. Banyak metode statistika dan machine learning non-generatif tidak menghasilkan bahasa atau media baru. Sebaliknya, LLM dapat membuat uraian atas hasil analisis tetapi belum tentu merupakan metode statistik yang valid. Peneliti sering memakai satu antarmuka untuk mengunggah data, menjalankan kode, dan menafsirkan output; hal itu menyatukan beberapa operasi yang secara ilmiah harus dibedakan. Misalnya, komputasi numerik mungkin dilakukan melalui Python di balik layar, sedangkan narasi disusun LLM. Keduanya memiliki sumber kesalahan dan bukti validasi yang berbeda.

Model harus dinilai dalam relasi terhadap tugas dan konteks penggunaan. Sistem yang layak untuk menghitung frekuensi kata pada korpus umum mungkin tidak layak untuk menilai risiko individu. Alat yang cukup membantu untuk membuat plot awal mungkin gagal menghasilkan kode khusus software

tertentu. Studi Prandner et al. (2025), yang mengevaluasi ChatGPT versi gratis dalam replikasi beberapa prosedur, menemukan tugas tertentu lebih dapat ditangani dibanding sintaks khusus yang lebih kompleks. Temuannya mengingatkan bahwa keberhasilan pada contoh sederhana tidak membuktikan generalitas pada semua tugas statistik.

3. Pertanyaan penelitian menentukan analisis

Analisis yang baik dimulai dengan memperjelas jenis pertanyaan. Apakah penelitian ingin menggambarkan distribusi, membandingkan kelompok, menguji hubungan, memperkirakan efek kausal, memprediksi outcome, memahami pengalaman, mengidentifikasi pola laten, atau mengevaluasi suatu sistem? Setiap tujuan mengarahkan konsep bukti yang berbeda. Prediksi yang akurat tidak otomatis menjelaskan sebab. Signifikansi statistik tidak otomatis menunjukkan dampak praktis. Tema yang sering muncul tidak otomatis tema yang paling penting. Ringkasan model tidak otomatis menjadi interpretasi.

Pada studi deskriptif, peneliti menentukan populasi, periode, denominator, unit analisis, dan definisi indikator. Pada studi inferensial, ia menetapkan estimand, mekanisme pemilihan sampel, asumsi pengujian, serta cara menangani ketidakpastian. Pada studi prediktif, ia menentukan target yang ingin diprediksi, waktu prediksi, biaya kesalahan, baseline, dan pembagian data yang mencerminkan penggunaan nyata. Pada studi kualitatif, pertanyaan menyangkut makna, pengalaman, proses, dan konteks yang perlu ditafsirkan. AI dapat membantu menghasilkan opsi analisis, tetapi tidak menentukan tujuan ilmiah tersebut.

Satu dataset dapat menjawab berbagai pertanyaan, tetapi tidak semua pertanyaan yang mungkin diajukan didukung oleh data. Misalnya data LMS bisa menunjukkan waktu login, pengiriman tugas, dan skor kuis. Data itu dapat menggambarkan pola keterlibatan yang terekam, tetapi tidak secara langsung mengukur motivasi, kualitas pemahaman, atau kondisi keluarga.

Menamai frekuensi klik sebagai “engagement” memerlukan teori dan validasi. Jika model AI menyebut variabel yang tersedia sebagai konstruk psikologis, peneliti harus menguji apakah pemetaan itu benar.

Sebelum memilih algoritma, tim menuliskan pertanyaan, populasi target, unit observasi, variabel hasil, prediktor, waktu pengukuran, dan konsekuensi keputusan. Langkah ini mencegah “method shopping”, yaitu mencoba banyak model sampai satu menghasilkan cerita yang disukai. Preregistrasi, rencana analisis, atau protokol kualitatif dapat membedakan prosedur yang direncanakan dari eksplorasi. AI dapat membantu mengkritik protokol dengan mengidentifikasi asumsi atau analisis alternatif, tetapi pilihan akhir dan alasan penerimaannya harus dicatat.

4. Peta siklus analisis data berbantuan AI

Alur kerja yang dapat dipercaya mencakup beberapa lapisan. Pertama, penetapan tujuan dan keputusan yang hendak didukung. Kedua, pemahaman sumber data serta hak penggunaannya. Ketiga, audit struktur dan kualitas. Keempat, transformasi atau pembersihan yang dapat dilacak. Kelima, analisis sesuai pertanyaan. Keenam, validasi teknis dan substantif. Ketujuh, interpretasi bersama konteks dan teori. Kedelapan, pelaporan, arsip, dan pengelolaan risiko pascapublikasi. AI dapat hadir pada satu atau beberapa lapisan, tetapi kehadiran tersebut harus terlihat dalam dokumentasi.

Pemetaan intervensi penting karena “human oversight” yang tidak spesifik mudah menjadi slogan. Untuk setiap penggunaan, peneliti menjawab: apa inputnya; operasi apa yang dilakukan sistem; apa outputnya; siapa memeriksa; standar apa yang dipakai; dan apa tindakan jika gagal? Misalnya untuk pengkodean jawaban terbuka, input adalah respons yang telah disamarkan, operasi adalah klasifikasi menurut codebook, output adalah label dan alasan, pemeriksaan dilakukan pada sampel oleh dua coder, dan konflik dibawa ke diskusi. Prosedur ini lebih terukur daripada kalimat “hasil AI telah diperiksa”.

Tidak semua operasi harus dilakukan dengan AI. Penghitungan rerata sederhana mungkin lebih transparan melalui fungsi statistik yang telah diuji. Pengodean tema laten mungkin membutuhkan interpretasi manusia. OCR dapat membantu ekstraksi dokumen, tetapi perlu koreksi pada bagian yang memengaruhi hasil. Keputusan menggunakan model dibuat per tugas. Jika metode konvensional lebih aman, efisien, dan cukup menjawab pertanyaan, tidak ada alasan metodologis untuk menambahkan model generatif.

Dalam pipeline bertingkat, kesalahan awal dapat menyebar. OCR yang salah mengubah kata; pembersihan otomatis menghapus istilah lokal; embedding mengelompokkan respons; LLM merangkum kelompok; peneliti kemudian menafsirkan ringkasan sebagai pola sosial. Tanpa memeriksa data pada setiap transformasi, kesalahan akhir sulit dilacak. Karena itu, pipeline menyimpan salinan data mentah, data antara, log transformasi, versi skrip, dan output model. Transformasi bersifat dapat dipulihkan bila memungkinkan.

5. Audit data sebelum analisis

Sebelum model dipilih, peneliti memeriksa asal data, izin penggunaan, skema variabel, satuan, rentang, kategori, tanggal, duplikasi, nilai hilang, kesalahan pencatatan, serta hubungan antarfile. Metadata menjelaskan siapa mengumpulkan data, kapan, dengan instrumen apa, tujuan awal, populasi, dan batas penggunaan ulang. Dataset dari platform digital sering terlihat terstruktur, tetapi dapat merekam perilaku hanya ketika pengguna memiliki akses dan memilih untuk berinteraksi. Tidak tercatat tidak sama dengan tidak terjadi.

Pemeriksaan awal membuat kamus data yang menjelaskan nama variabel, definisi, satuan, kode kategori, dan rentang yang sah. AI dapat membantu menghasilkan dokumentasi atau mendeteksi nilai yang tidak lazim, tetapi definisi dikonfirmasi terhadap formulir, buku kode, atau pembuat data. Nilai 999 bisa berarti missing code, bukan skor yang sangat tinggi. Tanggal dapat

terbaca sebagai teks; skor likert bisa disimpan dalam bentuk string; variabel kategori dapat kehilangan labelnya saat diekspor. Model tidak dapat menebak makna lokal secara andal tanpa dokumentasi.

Nilai hilang dianalisis menurut pola dan alasan yang mungkin. Menghapus semua baris yang memiliki missing value dapat mengurangi ukuran sampel dan mengubah komposisi. Mengisi setiap nilai dengan rerata dapat mengecilkan variasi dan merusak hubungan. Imputasi memerlukan asumsi yang dijelaskan dan analisis sensitivitas. Jika chatbot menyarankan “hapus data kosong”, peneliti menilai mekanisme missingness, proporsi, pola antarvariabel, serta dampak terhadap inferensi. Keputusan pembersihan disimpan agar hasil dapat diulang.

Pencilan juga bukan otomatis kesalahan. Nilai ekstrem mungkin mencerminkan kesalahan input, fenomena langka, atau subpopulasi penting. Aturan deteksi statistik perlu digabung dengan pemeriksaan sumber dan pengetahuan bidang. AI dapat menandai anomali tetapi tidak memutuskan penghapusan. Peneliti mencatat alasan koreksi atau retensi, menjalankan analisis dengan dan tanpa kasus berpengaruh jika relevan, lalu melaporkan apakah kesimpulan berubah.

Data dari berbagai sumber memerlukan pemeriksaan kesesuaian waktu, definisi, unit, dan populasi. Penggabungan data akademik, demografi, dan platform mungkin meningkatkan informasi, tetapi juga meningkatkan risiko re-identifikasi. Kunci penggabungan, resolusi entitas, dan tingkat kecocokan dicatat. AI dapat membantu fuzzy matching, namun false match dapat menyatukan rekam orang yang berbeda. Sampel audit manual dan ambang kecocokan disusun sebelum data dipakai untuk hasil utama.

6. Analisis deskriptif dan eksplorasi pola

Statistik deskriptif menjawab bagaimana data yang diamati tersusun: jumlah kasus, distribusi, pusat, variasi, frekuensi, dan pola waktu. AI dapat membantu

memilih visualisasi, menulis kode ringkas, atau menjelaskan tabel kepada pembaca. Namun, peneliti tetap memastikan denominator, label, unit, periode, dan skala benar. Rata-rata dapat menutupi distribusi miring; persentase tanpa jumlah dasar dapat menyesatkan; tren tidak mengungkap mekanisme penyebab. Grafik yang baik memperlihatkan struktur data tanpa mengklaim lebih dari yang dihitung.

Analisis eksploratori digunakan untuk mengenali anomali dan membangun hipotesis. Ia bukan konfirmasi otomatis. Jika model menyarankan ada kelompok pengguna aktif, peneliti memeriksa bagaimana “aktif” didefinisikan, apakah skala fitur masuk akal, dan apakah pengelompokan stabil pada variasi parameter. Clustering akan selalu membuat kelompok meskipun struktur populasi tidak memiliki klaster alami. Nilai silhouette atau ukuran pemisahan hanya menilai aspek tertentu; interpretasi substantif tetap memerlukan pemeriksaan.

AI dapat membuat visualisasi data lebih mudah dipahami bagi peneliti yang tidak akrab dengan bahasa pemrograman. Pengguna meminta diagram, kemudian memeriksa kode, tabel sumber, filter, agregasi, serta pengaturan sumbu. Sumbu dipotong dapat membesar-besarkan perbedaan; warna dapat tidak terbaca oleh pembaca dengan buta warna; garis tren dapat menyembunyikan variasi; dan label kategoris dapat salah diurutkan. Prinsip yang sama berlaku untuk grafik yang dibuat manual maupun otomatis. Hasil visual tidak diklaim sebagai bukti jika underlying computation belum diuji.

Eksplorasi harus meninggalkan jejak keputusan. Catat variabel yang dicoba, transformasi, filter, dan model eksploratif. Ini tidak berarti semua percobaan perlu dimasukkan ke artikel, tetapi analisis final harus membedakan hasil yang telah direncanakan dari temuan yang muncul sesudah melihat data. Banyaknya percobaan meningkatkan peluang pola kebetulan; AI

mempercepat pencarian pola sehingga kebutuhan koreksi dan replikasi justru meningkat.

7. Inferensi statistik dan pengujian hipotesis

Statistik inferensial menggunakan data sampel untuk memperkirakan atau menguji klaim tentang populasi di bawah asumsi tertentu. AI dapat membantu menulis kode uji-t, ANOVA, regresi, uji nonparametrik, atau interval kepercayaan. Akan tetapi, pemilihan uji harus mengikuti desain, skala, distribusi, ketergantungan observasi, dan pertanyaan. Dua kelompok independen berbeda dari pengukuran berpasangan; data bertingkat pada kelas dan sekolah memerlukan pertimbangan struktur klaster; outcome ordinal tidak selalu sesuai diperlakukan sebagai variabel kontinu tanpa justifikasi.

P-value bukan probabilitas bahwa hipotesis nol benar dan bukan ukuran besar efek. Model generatif dapat menghasilkan interpretasi yang terlalu sederhana seperti "$p<.05$ membuktikan variabel berpengaruh". Peneliti membaca koefisien, interval kepercayaan, ukuran efek, kualitas desain, dan relevansi praktis. Uji signifikansi yang sama dapat memberi nilai berbeda saat asumsi dilanggar, banyak pengujian dilakukan, atau ukuran sampel sangat besar. AI tidak menghapus persoalan inferensi; ia bisa mempercepat kesalahan jika peneliti hanya menyalin narasi.

Regresi membutuhkan pemeriksaan spesifikasi. Apakah variabel diukur secara tepat? Apakah hubungan linier? Apakah residual bermasalah? Apakah confounding ditangani sesuai tujuan? Apakah interaksi direncanakan? Apakah observasi independen? Model dapat menghasilkan sintaks yang benar secara komputasional tetapi salah secara substantif. Validasi kode berbeda dari validasi rancangan. Angka yang sama dengan software lain tidak membuktikan bahwa model cocok dengan pertanyaan atau asumsi penelitian.

Dalam model SEM atau PLS-SEM, penggunaan AI untuk menghitung indikator atau membuat interpretasi harus mengikuti prosedur pengukuran dan struktur

model. Peneliti menilai reflektif atau formatif berdasarkan teori, kualitas item, reliabilitas, validitas konvergen dan diskriminan, collinearity, serta kesesuaian model. AI dapat membantu mengaudit perhitungan atau menghasilkan tabel, tetapi tidak membenarkan penggunaan ambang secara mekanis. Nilai loading atau HTMT perlu ditafsirkan bersama konteks konstruk dan rancangan. Saran model bukan pengganti literatur metodologis.

Pada analisis kelompok, multiple testing, atau subgroup analysis, keputusan perlu dijelaskan. Eksplorasi banyak subgroup dapat menghasilkan perbedaan kebetulan. Peneliti menetapkan analisis utama sebelumnya jika mungkin, menggunakan koreksi yang sesuai, dan melaporkan hasil eksploratori sebagai eksploratori. AI yang menyarankan subgroup berdasarkan hasil akhir dapat memperbesar bias konfirmasi. Transparansi membantu pembaca membedakan bukti yang direncanakan dan hipotesis baru.

8. Machine learning untuk prediksi dan klasifikasi

Machine learning berorientasi pada kinerja untuk tugas tertentu, misalnya memprediksi kelulusan, mengklasifikasikan kategori dokumen, atau mendeteksi pola anomali. Sebelum melatih model, peneliti mendefinisikan target, horizon waktu, populasi yang akan menerima prediksi, baseline pembanding, dan biaya kesalahan. Target yang mudah tersedia tidak selalu bermakna. Prediksi nilai akhir dari data selama semester dapat membantu dukungan akademik, tetapi dapat juga mengulang ketimpangan kesempatan jika variabel prediktor menangkap akses perangkat, disabilitas, atau kondisi sosial.

Pembagian data adalah keputusan inti. Train, validation, dan test split harus mencegah informasi dari masa depan atau dari individu yang sama masuk ke sisi uji. Pada data longitudinal, pembagian acak baris dapat menyebabkan kebocoran jika rekam mahasiswa yang sama tersebar di train dan test. Pada prediksi masa depan, validasi temporal lebih sesuai: latih pada periode

terdahulu dan uji pada periode berikutnya. Jika model menerima fitur yang baru diketahui setelah outcome terjadi, akurasi menjadi ilusi. Data leakage telah dikenali sebagai sumber kesalahan serius dalam riset machine learning berbasis sains (Kapoor & Narayanan, 2023).

Ukuran kinerja dipilih sesuai distribusi kelas dan tujuan. Akurasi dapat terlihat tinggi jika hampir semua kasus berada pada satu kelas. Precision dan recall menjawab konsekuensi berbeda; ROC-AUC dan PR-AUC memiliki interpretasi berlainan; kalibrasi menilai apakah probabilitas prediksi sesuai frekuensi kejadian. Untuk keputusan yang berdampak pada individu, evaluasi juga mengukur false positive, false negative, subgroup performance, dan konsekuensi keputusan. Model yang unggul pada rata-rata dapat gagal pada kelompok kecil.

Model kompleks tidak selalu unggul dalam konteks nyata. Bandingkan dengan baseline yang sederhana dan mudah ditafsirkan. Lakukan validasi eksternal pada institusi atau periode yang tidak dipakai untuk pelatihan, sebab distribusi data dapat bergeser. Perubahan kurikulum, kebijakan, sistem LMS, dan populasi mahasiswa dapat menyebabkan model kehilangan kalibrasi. Jika kinerja turun, institusi perlu mengetahui kapan model harus dikalibrasi ulang atau dihentikan.

ChatGPT atau LLM dapat membantu membangun kode machine learning, namun kode otomatis perlu diperiksa terhadap leakage, preprocessing, split, pemilihan fitur, tuning, dan pelaporan. Seluruh preprocessing yang mempelajari parameter dari data harus dipasang hanya pada data latih sebelum diterapkan ke validation atau test. Jika imputasi atau normalisasi dihitung sebelum split, informasi test bocor ke model. Peneliti menjalankan pipeline dalam kerangka yang dapat diulang dan menyimpan seed, dependensi, dan versi model.

9. Analitik kausal dan batas prediksi

Prediksi menjawab "siapa atau apa yang kemungkinan mengalami outcome?" Analisis kausal bertanya "apa perubahan yang akan terjadi jika intervensi berbeda?" Kedua tujuan tersebut berbeda. Algoritma prediktif dapat memanfaatkan korelasi yang tidak relevan secara kausal. Jika siswa yang lebih sering login memiliki nilai lebih tinggi, memberi tugas login tambahan tidak otomatis menaikkan nilai. Hubungan dapat dijelaskan oleh motivasi, akses, pengetahuan awal, atau faktor lain.

Estimasi kausal membutuhkan desain dan asumsi identifikasi: eksperimen acak, natural experiment, difference-in-differences, instrumental variables, regression discontinuity, matching, atau strategi lain yang sesuai. AI dapat membantu memetakan variabel potensial atau menguji spesifikasi, tetapi tidak menjamin confounding hilang. Pemilihan kovariat berdasarkan korelasi semata dapat menimbulkan bias, misalnya mengontrol mediator atau collider. Peneliti menggambar struktur kausal dan menjelaskan alasan tiap variabel dimasukkan.

Model bahasa sering menyusun narasi sebab-akibat dari tabel korelasi karena bahasa sehari-hari menggunakan "mempengaruhi" secara longgar. Peneliti menghindari mengadopsi kata kerja kausal bila desain tidak mendukungnya. Hasil observasional dapat menunjukkan asosiasi, hubungan kondisional, atau prediksi, tetapi kausalitas memerlukan argumen tambahan. Dalam evaluasi program tutoring berbasis AI, perbedaan nilai antara pengguna dan nonpengguna dapat mencerminkan seleksi diri. Untuk mengklaim efek program, kelompok pembanding dan proses penempatan perlu dirancang.

Estimasi efek heterogen dapat membantu memahami siapa yang memperoleh manfaat, tetapi analisis subgroup harus direncanakan dan divalidasi. Model causal forest atau meta-learner dapat menghasilkan efek individual, namun hasilnya sensitif pada data, confounding, dan pilihan algoritma. Prediksi efek

bukan fakta personal yang pasti. Jika digunakan untuk mengalokasikan sumber daya, keputusan perlu mempertimbangkan ketidakpastian, fairness, serta dukungan bagi kelompok yang salah diprediksi.

10. Analisis kualitatif, teks, dan bahasa alami

Data kualitatif dapat berupa wawancara, diskusi kelompok, jawaban terbuka, dokumen kebijakan, catatan observasi, surel, atau unggahan digital. AI menawarkan transkripsi, terjemahan, pencarian entitas, analisis sentimen, clustering, coding awal, dan ringkasan. Manfaatnya terutama pada pengelolaan korpus dan eksplorasi. Namun, pemaknaan bukan sekadar frekuensi kata. Ironi, metafora, jeda, konteks budaya, relasi kuasa, dan makna implisit dapat hilang dari ringkasan otomatis.

Peneliti dapat memakai model untuk mencari kutipan yang berkaitan dengan kode, menyarankan subkategori, atau membandingkan kasus. Kode awal harus ditautkan ke segmen asli dan diperiksa oleh peneliti. Codebook memuat definisi, kriteria inklusi-eksklusi, contoh, serta aturan kasus ambigu. Jika dua coder berbeda pendapat, perbedaan itu dibahas; model tidak dijadikan pemutus karena outputnya tampak konsisten. Studi tentang penggunaan LLM dalam analisis kualitatif menekankan potensi bantuan bersama kebutuhan menjaga konteks, refleksivitas, dan akuntabilitas manusia (Jalali & Akhavan, 2024; Nguyen-Trung, 2025).

Analisis sentimen memerlukan pemeriksaan bahasa dan konteks. Kata positif dapat memiliki makna negatif secara sarkastik; ungkapan sopan dapat menyembunyikan kritik; istilah lokal dapat salah diklasifikasi. Model yang dilatih terutama pada bahasa dominan mungkin tidak stabil pada bahasa Indonesia informal, campur kode, atau bahasa daerah. Peneliti membuat set validasi lokal dan mengecek false positive serta false negative pada subkelompok yang berbeda.

Topic modeling dan embedding membantu menemukan kemiripan dokumen, tetapi jarak vektor bukan makna substantif. Algoritma mengelompokkan berdasarkan representasi yang dipelajari; peneliti menentukan apakah hasil klaster koheren dan berguna terhadap pertanyaan. Jumlah klaster, preprocessing, model embedding, dan random seed dapat mengubah hasil. Validasi memadukan ukuran komputasional dan penilaian manusia, serta menyimpan keputusan tentang label tema.

Dalam penelitian interpretif, posisi peneliti dan proses pembentukan tema adalah bagian metodologi. Mengotomatisasi coding secara penuh dapat mengubah asumsi epistemologis studi. Jika penelitian menggunakan reflexive thematic analysis, misalnya, kesepakatan antar-coder tidak selalu menjadi kriteria mutu utama. AI dapat membantu eksplorasi, tetapi peneliti menjelaskan bagaimana ia memengaruhi keterlibatan dengan data dan apakah sistem ditempatkan sebagai alat clerical, partner dialogis, atau coder awal.

11. Metode campuran dan integrasi hasil

Metode campuran mengintegrasikan bukti numerik dan naratif untuk menjawab persoalan kompleks. AI dapat membantu merangkum dataset, menyusun joint display, atau menandai temuan yang selaras dan tidak selaras. Integrasi tidak terjadi otomatis hanya karena dua tabel dimasukkan ke chatbot. Rancangan perlu menentukan urutan fase, prioritas, titik penghubung, serta logika meta-inferensi. Combrinck (2024) menunjukkan penggunaan GenAI pada analisis dan integrasi mixed methods, tetapi tutorial berbasis contoh tertentu tidak menggantikan justifikasi desain pada studi lain.

Contoh: survei menunjukkan mahasiswa yang menggunakan AI lebih sering menyelesaikan tugas tepat waktu; wawancara menemukan sebagian mahasiswa menghabiskan waktu lebih banyak untuk memeriksa jawaban. AI dapat menyarankan narasi bahwa AI “meningkatkan efisiensi dengan risiko

tertentu". Peneliti melangkah lebih jauh dengan membandingkan jenis tugas, pengalaman, akses, kualitas output, dan kelompok peserta. Perbedaan bukti dapat menunjukkan bahwa efisiensi meningkat pada tahap drafting tetapi waktu validasi bertambah pada tugas analitik.

Joint display mempertahankan sumber: kolom kuantitatif berisi estimasi dan ketidakpastian; kolom kualitatif berisi kutipan yang diverifikasi; kolom integrasi menjelaskan hubungan bukti. AI tidak membuat kutipan, angka, atau hubungan yang tidak tersedia di data. Temuan konvergen, komplementer, dan divergen dilaporkan terpisah. Model dapat membantu menata, tetapi peneliti bertanggung jawab atas meta-inferensi serta pertimbangan konteks.

12. Analisis data akademik dan learning analytics

Institusi pendidikan mengumpulkan data administratif, nilai, LMS, perpustakaan, kehadiran, interaksi digital, survei, serta komunikasi pembimbing. Analitik pembelajaran dapat membantu mengenali mahasiswa yang membutuhkan dukungan, memahami pola partisipasi, dan mengevaluasi desain kurikulum. Namun, jejak digital bukan cerminan utuh pembelajaran. Klik dapat menunjukkan akses, bukan perhatian; waktu pada halaman dapat menunjukkan tab terbuka, bukan membaca; nilai kuis dapat mencerminkan pengetahuan, strategi menebak, atau kualitas soal.

Model prediksi risiko akademik dapat mendukung intervensi jika tujuannya jelas dan mahasiswa memperoleh manfaat. Penggunaan skor untuk melabeli mahasiswa sebagai "berisiko" tanpa dukungan dapat menimbulkan stigma, pengawasan, atau pengurangan kesempatan. Threshold harus dirancang menurut kapasitas intervensi, biaya kesalahan, dan hak untuk meninjau keputusan. Dosen perlu memahami bahwa prediksi bersifat probabilistik dan dapat salah, bukan diagnosis permanen tentang kemampuan.

Learning analytics memerlukan keterlibatan pihak yang terkena dampak. Mahasiswa perlu mengetahui jenis data yang digunakan, tujuan, siapa yang

mengakses, berapa lama disimpan, dan apakah hasil digunakan untuk evaluasi atau dukungan. Tata kelola memperjelas persetujuan, akses, retensi, mekanisme koreksi, dan batas penggunaan sekunder. Literatur human-centred learning analytics mengingatkan bahwa pertumbuhan sistem data intensif berkaitan dengan isu privasi dan agensi peserta (Alfredo et al., 2024).

Kualitas sistem diukur bukan hanya melalui akurasi prediksi, tetapi juga apakah intervensi efektif, dapat diakses, tidak menambah kesenjangan, dan tidak menimbulkan beban yang tidak proporsional. Institusi melakukan pilot, mengevaluasi dampak kausal jika memungkinkan, mengkaji kinerja lintas program dan kelompok, dan menghentikan sistem jika tidak bermanfaat. Data akademik adalah aset untuk pembelajaran dan perbaikan, bukan sekadar bahan pengawasan administratif.

13. Analitik penelitian manajemen dan organisasi

Dalam manajemen, AI dapat menganalisis survei pegawai, data penjualan, ulasan pelanggan, dokumen kebijakan, jaringan kolaborasi, indikator operasional, dan data platform kerja. Pemetaan pola dapat membantu memahami pengalaman karyawan, efektivitas program pelatihan, adopsi teknologi, atau hubungan strategi dengan kinerja. Namun, indikator bisnis juga dipengaruhi definisi akuntansi, insentif, musim, pergantian kebijakan, dan kualitas sistem pencatatan. Menyatukan data dari unit berbeda dapat menghasilkan perbandingan yang tidak setara.

Misalnya, tim mengevaluasi pelatihan digital dan menemukan karyawan yang menyelesaikan lebih banyak modul memiliki kinerja lebih tinggi. Kesimpulan “pelatihan meningkatkan kinerja” belum pasti karena pegawai berkinerja tinggi mungkin lebih punya waktu dan motivasi mengikuti modul. Desain memperhitungkan baseline, seleksi peserta, supervisi, jenis pekerjaan, dan periode outcome. Jika tidak ada randomisasi, analisis menjelaskan batas atribusi kausal.

Ulasan pelanggan dapat diringkas dengan NLP untuk mengidentifikasi isu layanan. Namun, pelanggan yang memberi ulasan bukan sampel acak, platform mungkin menekan jenis komentar tertentu, dan sentimen berbeda menurut gaya bahasa. Model dapat salah mengerti sarkasme atau campur bahasa. Organisasi sebaiknya memakai hasil sebagai sinyal untuk investigasi, bukan sebagai representasi pasti semua pelanggan atau dasar tunggal keputusan karyawan.

Analisis jaringan organisasi memerlukan definisi relasi: kolaborasi proyek, konsultasi, komunikasi, atau persetujuan. Matriks jaringan berdasarkan metadata surel dapat mengungkap pola struktural, tetapi penggunaan harus menghormati privasi dan konteks pekerjaan. Sentralitas tidak otomatis berarti pengaruh, dan jaringan komunikasi bukan pengukuran langsung kualitas kepemimpinan. Interpretasi ditinjau bersama pemilik proses dan prinsip minimalisasi data.

14. Validasi dan replikasi analisis

Validasi bertanya apakah prosedur melakukan tugas yang dimaksud pada data dan konteks yang tepat. Dalam klasifikasi, pembanding dapat berupa label ahli dengan pedoman yang jelas. Dalam prediksi, model diuji pada data yang tidak dipakai melatih dan, jika memungkinkan, institusi atau waktu berbeda. Dalam kode statistik, hasil dibandingkan dengan output yang diketahui, software lain, atau perhitungan manual pada contoh kecil. Dalam coding kualitatif, kutipan dan codebook ditinjau oleh peneliti yang kompeten. Validasi tidak tunggal; setiap metode memerlukan bentuk pembuktian berbeda.

Replikasi memerlukan informasi tentang pipeline: sumber data, filter, transformasi, definisi outcome, versi kode, pustaka, parameter, random seed, lingkungan, dan aturan penghapusan. Jika LLM menulis sebagian skrip, arsipkan prompt utama, model dan tanggal penggunaan, output yang dipertahankan, serta revisi manusia. Untuk penggunaan layanan awan yang

berubah, keluaran tidak selalu identik; simpan output yang digunakan dalam analisis dan jelaskan keterbatasan mengulang panggilan model.

Reproducibility juga memerlukan pemisahan antara artifact yang tahan lama dan pemanggilan model yang berubah. Bila mungkin, tempatkan LLM "off the data path": gunakan untuk menghasilkan skrip atau dokumentasi yang kemudian diperiksa dan disimpan, bukan menjalankan transformasi penting secara tersembunyi pada setiap pengulangan. Jika model tetap menjadi bagian hidup pipeline, dokumentasikan determinisme, perubahan versi, serta hasil sensitivitas. Literatur machine learning menekankan masalah leakage, kualitas data, metrik, dan ketersediaan kode sebagai ancaman terhadap reproduksibilitas (Kapoor & Narayanan, 2023).

Peneliti membedakan reproducibility (hasil dapat dihasilkan kembali dari data dan prosedur yang sama), replicability (temuan diuji menggunakan data baru atau kondisi baru), dan robustness (kesimpulan bertahan pada pilihan analisis yang masuk akal). AI dapat mempermudah re-run tetapi tidak otomatis memberi replikasi eksternal. Pemeriksaan robustness dilakukan dengan spesifikasi alternatif yang relevan, bukan mencoba banyak opsi sampai hasil signifikan.

15. Bias, fairness, dan kualitas generalisasi

Bias dapat masuk pada setiap tahap: populasi yang dipilih, alat ukur, proses pencatatan, label outcome, data pelatihan, algoritma, threshold, dan cara hasil digunakan. Dataset besar tidak menjamin representasi adil; model dapat sangat tepat untuk kelompok mayoritas tetapi salah bagi kelompok kecil. Penggunaan skor agregat menutupi ketimpangan ini. Audit mempertimbangkan kelompok yang relevan secara etis, ukuran sampel masing-masing, serta ketidakpastian estimasi.

Keadilan bukan hanya persoalan angka metrik. Model yang sama dapat menimbulkan dampak berbeda tergantung proses organisasi. Skor risiko

akademik mungkin mendorong bantuan lebih awal jika disertai dukungan; ia bisa membatasi akses jika dipakai untuk menolak mahasiswa. Model seleksi karyawan mungkin mengulang pola historis diskriminatif jika label "kinerja" memuat keputusan lama. Obermeyer et al. (2019) menunjukkan bagaimana penggunaan biaya kesehatan sebagai proksi kebutuhan dapat menciptakan bias rasial dalam algoritma manajemen kesehatan, mengingatkan bahwa kualitas label dan proksi sama pentingnya dengan algoritma.

Generalisasi dibatasi perubahan waktu, teknologi, budaya, bahasa, kurikulum, ekonomi, dan kebijakan. Model yang dilatih di satu universitas tidak otomatis berlaku di universitas lain. Distribusi prediktor dan hubungan prediktor-outcome dapat bergeser. Validasi lokal bukan sekadar formalitas, terutama ketika sistem digunakan untuk keputusan yang memengaruhi orang. Peneliti melaporkan siapa yang masuk dan tidak masuk sampel serta kemungkinan konsekuensi untuk kelompok yang kurang terlihat.

16. Explainability dan komunikasi ketidakpastian

Explainability mencakup beberapa pertanyaan: bagaimana model menghasilkan prediksi, fitur apa yang berkontribusi, apakah penjelasan stabil, dan apakah penjelasan membantu pengguna membuat keputusan yang lebih baik? Importance score atau heatmap bukan penjelasan sebab-akibat. Penjelasan post hoc dapat menyederhanakan mekanisme model dan berubah ketika data atau sampel sedikit berbeda. Peneliti menjelaskan tujuan interpretasi dan menguji stabilitas penjelasan.

Untuk model yang digunakan mendukung kebijakan, pembaca perlu memahami faktor yang mendorong prediksi, rentang ketidakpastian, batas populasi, dan tindakan yang dapat diambil. Penjelasan kepada pengguna harus sesuai pemahaman dan hak mereka. Model dapat secara statistik menemukan bahwa keterlambatan tugas berkorelasi dengan risiko putus studi, tetapi interpretasi institusi mempertimbangkan hambatan akses, beban

kerja, atau tanggung jawab keluarga. Penjelasan yang bermanfaat menghubungkan komputasi dengan mekanisme yang mungkin dan bukti tambahan.

Ketidakpastian muncul dari sampling, pengukuran, spesifikasi model, perubahan populasi, dan ketidakstabilan sistem. Laporan menghindari angka probabilitas tunggal tanpa kalibrasi atau interval. Prediksi 0,72 bukan jaminan individu akan memperoleh hasil tertentu. Dalam komunikasi publik, visualisasi ketidakpastian perlu membantu pembaca menghindari kepastian palsu. Model generatif dapat menulis kalimat yang tegas; peneliti mengoreksi tingkat kepastian agar sepadan dengan bukti.

17. Privasi, persetujuan, dan tata kelola data

Dataset akademik dapat mengandung identitas langsung, atribut sensitif, atau kombinasi ciri yang memungkinkan identifikasi ulang. Menghapus nama saja tidak selalu cukup. Peneliti menilai risiko sebelum mengunggah data ke platform AI, membaca kebijakan penyimpanan dan pelatihan, serta memahami lokasi pemrosesan. Data sensitif tidak dimasukkan ke layanan konsumen tanpa dasar persetujuan dan perlindungan. Pilihan dapat berupa model lokal, lingkungan komputasi institusional, data sintetis untuk pengujian pipeline, atau pengolahan data minimal.

Persetujuan partisipan menjelaskan tujuan data, akses, penggunaan sekunder, pemrosesan oleh vendor, retensi, dan hak yang berlaku. Penelitian pendidikan dan organisasi memiliki relasi kuasa: mahasiswa mungkin merasa wajib berpartisipasi; pegawai mungkin khawatir jawaban diketahui atasan. Pemisahan identitas, kontrol akses, dan pelaporan agregat perlu mengurangi risiko. Komite etik menilai apakah penggunaan AI mengubah tingkat risiko atau proses interaksi.

Tata kelola menetapkan penanggung jawab data, dasar akses, jadwal retensi dan penghapusan, penanganan insiden, dan pemeriksaan vendor. Dataset

dipakai sesuai tujuan yang disetujui, bukan otomatis untuk eksperimen lain karena model tersedia. Model yang memproses data institusi harus mengikuti kebijakan perlindungan data dan keamanan informasi. Artikel ini memberikan prinsip metodologis umum; aturan hukum dan kebijakan sektor harus diperiksa melalui sumber resmi terkini.

18. Synthetic data dan simulasi

Data sintetis dihasilkan untuk meniru sifat data tanpa menyalin seluruh rekaman asli. Ia dapat digunakan menguji perangkat lunak, membuat contoh pengajaran, atau menguji alur analitik ketika akses data nyata dibatasi. Namun, kesamaan rerata dan korelasi sederhana tidak membuktikan bahwa data sintetis mempertahankan struktur, distribusi, relasi antarvariabel, subgroup, atau mekanisme kausal. Peneliti menjelaskan generator, tujuan, metrik utility, risiko privacy leakage, dan perbedaan yang ditemukan.

Synthetic records tidak menjadi responden empiris. Model bahasa dapat membuat jawaban survei atau wawancara yang terdengar realistis, tetapi distribusinya mencerminkan pola dalam data pelatihan dan prompt, bukan pengalaman populasi target. Menggunakan data buatan untuk memperkirakan preferensi konsumen atau sikap mahasiswa tanpa validasi eksternal akan menghasilkan klaim yang lemah. Data sintetis paling aman dipakai sebagai alat teknis atau sumber hipotesis, bukan pengganti otomatis bukti manusia.

Simulasi agen dapat membantu mengeksplorasi skenario interaksi atau kebijakan, tetapi perilaku agen bergantung pada deskripsi persona, prompt, model, dan aturan interaksi. Peneliti melaporkan konfigurasi serta menguji apakah hasil stabil pada model atau prompt lain. Simulasi memperluas ruang eksplorasi, bukan membuat dunia sosial tersedia secara langsung di dalam model.

19. Visualisasi dan narasi hasil dengan bantuan AI

Visualisasi membantu melihat pola, membandingkan kelompok, dan menyampaikan temuan. Model dapat menulis kode chart atau menawarkan bentuk grafik, tetapi pilihan visualisasi mengikuti pertanyaan dan tipe variabel. Diagram batang sesuai untuk perbandingan kategori, histogram dan density plot untuk distribusi, scatterplot untuk hubungan dua variabel, dan peta untuk data spasial jika skala serta lokasi dapat ditampilkan secara etis. Visualisasi tidak dipilih hanya karena tampilannya menarik.

Grafik AI-generated diperiksa pada tabel dasar: jumlah kasus, penyebut persentase, transformasi, filter, sumbu, urutan kategori, dan label. Warna, skala, dan anotasi harus meminimalkan bias perseptual. Grafik perlu membedakan data mentah, prediksi, interval, dan imputasi. Jika model menyarankan caption, peneliti memeriksa apakah caption sesuai hasil statistik dan tidak mengubah asosiasi menjadi klaim kausal.

Narasi hasil dapat dibantu AI untuk merangkum tabel dan menyesuaikan bahasa pembaca. Namun, narasi model dapat melewatkan hasil nol, menyebut arah efek secara keliru, atau menambahkan penjelasan yang tidak diuji. Peneliti melakukan pemeriksaan baris demi baris terhadap tabel, model, dan pertanyaan penelitian. Pernyataan hasil membedakan temuan utama, analisis eksplorasi, keterbatasan, dan interpretasi. Keterbacaan penting, tetapi tidak boleh dibeli dengan menghilangkan ketidakpastian.

20. Workflow analisis yang dapat direplikasi

Langkah pertama ialah menulis tujuan dan estimand secara sederhana. Misalnya, “membandingkan perubahan skor literasi digital antara mahasiswa yang ditawari lokakarya dan mahasiswa kelompok pembanding setelah satu semester,” bukan “melihat pengaruh AI”. Tujuan menjadi kriteria untuk memilih outcome, populasi, pembanding, dan horizon waktu. Tim juga menentukan apa yang tidak dapat disimpulkan.

Langkah kedua ialah membuat data inventory dan kamus variabel, memeriksa izin, serta menyimpan salinan data mentah. Langkah ketiga ialah menetapkan aturan pembersihan sebelum hasil final diketahui. Keputusan nilai hilang, duplikasi, outlier, dan penggabungan dicatat. Langkah keempat adalah memilih metode dan baseline. Bila model prediktif dipakai, pembagian data dan metrik ditentukan sebelum tuning. Bila analisis kualitatif dilakukan, codebook dan aturan audit disepakati.

Langkah kelima ialah memakai AI untuk tugas yang spesifik. Prompt menyertakan konteks secukupnya, mendefinisikan output yang diinginkan, menyebut batasan, dan meminta transparansi tentang asumsi. Prompt tidak memuat informasi rahasia yang tak diperlukan. Keluaran disimpan bersama waktu, model, versi, dan perubahan manual. Jika model mengusulkan kode, peneliti menjalankannya di salinan lingkungan aman dan memeriksa hasil pada contoh sederhana.

Langkah keenam ialah validasi. Tim membandingkan hasil dengan perhitungan independen, baseline, penilaian ahli, atau sumber primer sesuai jenis tugas. Langkah ketujuh ialah analisis sensitivitas, subgroup, dan robustness yang relevan. Langkah kedelapan ialah interpretasi dengan teori dan konteks, lalu pelaporan semua keputusan substantif. Langkah kesembilan ialah arsip kode, dokumentasi, bahan yang dapat dibagi, disclosure AI, serta jadwal peninjauan bila model dipakai dalam operasional.

21. Contoh kasus pendidikan tinggi

Sebuah universitas ingin mengidentifikasi mata kuliah dengan risiko kegagalan tinggi untuk memberikan dukungan lebih awal. Dataset mencakup nilai kuis, kehadiran, waktu pengumpulan, interaksi LMS, dan survei awal. Sebelum melatih model, universitas mendefinisikan tujuan: memprioritaskan undangan tutoring sukarela, bukan menentukan status akademik. Outcome adalah kegagalan pada akhir semester; waktu prediksi ditetapkan setelah minggu

keempat agar intervensi masih mungkin dilakukan. Data yang terjadi setelah minggu keempat tidak digunakan.

Tim membagi data berdasarkan angkatan sehingga test set merepresentasikan cohort masa depan. Baseline sederhana, seperti skor kuis awal dan kehadiran, dibandingkan dengan random forest dan gradient boosting. Evaluasi melaporkan recall, precision, calibration, dan jumlah mahasiswa yang dapat ditangani oleh kapasitas tutor. Kinerja diperiksa menurut program dan kelompok relevan dengan batasan privasi. Fitur yang merepresentasikan perangkat mahal tidak otomatis dipakai karena dapat menjadi proksi kondisi ekonomi.

Sebelum implementasi, dosen dan mahasiswa diajak meninjau alur. Skor tidak dikirim ke dosen sebagai label tetap; dashboard memberi rentang ketidakpastian dan menyarankan percakapan dukungan. Mahasiswa dapat meminta koreksi jika data salah. Pilot dibandingkan dengan layanan biasa untuk menilai apakah undangan meningkatkan partisipasi dan keberhasilan, bukan hanya apakah prediksi akurat. Jika sistem menandai banyak mahasiswa tetapi dukungan tidak tersedia, penggunaannya perlu ditunda.

GenAI dapat membantu membuat dokumentasi model, menjelaskan variabel, atau menyusun ringkasan cohort. Data individu tidak diunggah ke chatbot publik. Sebelum laporan diterbitkan, angka diverifikasi dan batas kausal dijelaskan. Kasus ini menunjukkan bahwa mutu analisis bergantung pada hubungan antara prediksi, kapasitas intervensi, hak mahasiswa, dan evaluasi dampak.

22. Contoh kasus manajemen sumber daya manusia

Organisasi menggunakan analitik untuk memahami hubungan pelatihan, keterlibatan, dan kinerja. Data menggabungkan sistem HR, survei, catatan pembelajaran digital, dan evaluasi supervisor. Sebelum menganalisis, tim memeriksa konsistensi definisi jabatan dan periode, alasan data hilang, serta

akses yang sah. Identitas karyawan dipisahkan; hasil dilaporkan pada tingkat kelompok dengan ukuran minimum untuk mengurangi identifikasi.

Pertanyaan awal mungkin bersifat deskriptif: kelompok mana yang memiliki akses pelatihan? Pertanyaan evaluatif menilai apakah program menghasilkan peningkatan keterampilan. Pertanyaan prediktif memperkirakan kebutuhan dukungan. Tiap tujuan memerlukan desain berbeda. Jika peserta memilih pelatihan sendiri, perbandingan langsung dengan nonpeserta akan bias. Peneliti dapat menggunakan matching atau desain lain yang sesuai, tetapi menjelaskan asumsi dan melakukan sensitivitas terhadap confounding.

Model bahasa dapat membantu mengelompokkan komentar survei, tetapi komentar negatif tentang atasan adalah data sensitif. Sistem tidak digunakan untuk menilai individu atau menentukan promosi. Peneliti memeriksa interpretasi tema bersama ahli HR dan memastikan kutipan tidak dapat mengidentifikasi responden. Prediksi yang dihasilkan dipakai untuk perbaikan program agregat, bukan profil personal tanpa persetujuan.

Penggunaan hasil dilakukan melalui tata kelola lintas fungsi. HR, tim data, komite etik, dan perwakilan karyawan membahas tujuan, risiko, dan jalur keberatan. Organisasi mengevaluasi dampak pada kelompok berbeda dan tidak menganggap korelasi antara pelatihan dan kinerja sebagai bukti manfaat program. Studi kasus ini memperlihatkan mengapa analisis data harus mencakup implikasi keputusan, bukan hanya performa algoritma.

23. Peran kompetensi manusia dan kolaborasi tim

Peneliti yang memakai AI perlu memiliki kemampuan membaca data, statistik, teori, dan konteks bidang. Ia tidak harus menulis seluruh kode dari ingatan, tetapi harus mampu menjelaskan apa yang dilakukan kode dan memeriksa keluaran. Jika pengguna tidak dapat menilai apakah uji sesuai pertanyaan, kode otomatis tidak aman untuk menjadi otoritas. Kompetensi domain

diperlukan untuk mengenali variabel yang salah label, hasil yang mustahil, atau inferensi yang tidak sejalan dengan praktik nyata.

Tim analitik idealnya menggabungkan keahlian substansi, metodologi, statistik, manajemen data, perlindungan privasi, dan komunikasi. Peran dibagi dengan jelas: siapa mengelola data, siapa menulis kode, siapa memvalidasi, siapa membuat keputusan, dan siapa menyetujui publikasi. AI dapat mengurangi pekerjaan rutin, tetapi juga memindahkan beban audit kepada peneliti yang mungkin tidak punya wewenang atau waktu. Institusi menyediakan dukungan, bukan sekadar lisensi perangkat lunak.

Literasi AI mencakup pemahaman probabilistik, keterbatasan data pelatihan, bias, prompt sensitivity, versi, dan model drift. Literasi statistik mencakup perbedaan deskripsi, prediksi, inferensi, dan kausalitas. Literasi etik mencakup persetujuan, minimisasi data, fairness, kerahasiaan, serta dampak keputusan. Pelatihan terbaik menggunakan kasus nyata: kode salah, leakage, sitasi palsu, atau ringkasan yang menghilangkan kutipan penentang. Dengan demikian, peneliti belajar dari kegagalan dan bukan hanya demonstrasi sukses.

24. Tata kelola institusi dan kebijakan data

Institusi membuat klasifikasi risiko untuk membedakan penggunaan rendah, sedang, dan tinggi. Penyuntingan tata bahasa pada naskah tanpa data sensitif berisiko berbeda dari pemrosesan transkrip pasien atau prediksi kinerja individu. Setiap kategori menetapkan jenis persetujuan, layanan yang boleh dipakai, validasi minimum, dan dokumentasi. Kebijakan cukup jelas untuk dipraktikkan tetapi ditinjau secara berkala karena fitur dan persyaratan vendor berubah.

Infrastruktur yang baik mencakup repositori aman, layanan AI yang ditinjau, akun institusional, logging yang proporsional, dan bantuan teknis. Institusi memperjelas retensi data, penggunaan input untuk pelatihan, lokasi penyimpanan, akses vendor, dan prosedur penghapusan. Peneliti tidak

seharusnya menafsirkan diamnya kebijakan sebagai izin mengunggah data rahasia.

Komite etik, petugas perlindungan data, perpustakaan, dan unit riset bekerja lintas fungsi. Komite etik menilai interaksi dan hak subjek; unit data menilai keamanan; pustakawan membantu sumber dan referensi; tim statistik mendukung validasi. Kebijakan authorship menjelaskan bahwa alat AI bukan penulis dan manusia bertanggung jawab atas konten. Peer reviewer juga menjaga kerahasiaan naskah dan mematuhi pedoman penerbit.

Institusi mengukur keberhasilan melalui kualitas penelitian, reproduksibilitas, dampak bagi peserta, akses yang adil, dan pembelajaran organisasi. Jumlah penggunaan AI bukan indikator kualitas. Aturan yang memberi insentif pada jumlah output dapat mendorong otomatisasi tanpa validasi. Laporan insiden dan kesalahan digunakan untuk memperbaiki sistem, bukan disembunyikan karena takut sanksi.

25. Keterbatasan, biaya, dan dampak ekologis

AI memiliki biaya komputasi, lisensi, tenaga validasi, penyimpanan, integrasi, dan ketergantungan vendor. Manfaat efisiensi dinilai terhadap keseluruhan pipeline: persiapan, prompt, pemeriksaan, koreksi, dan penanganan kesalahan. Tugas yang menghasilkan output instan tetapi membutuhkan audit panjang belum tentu menghemat waktu. Institusi juga memeriksa apakah platform akan tetap tersedia, dapat mengekspor data, atau mengubah model tanpa pemberitahuan.

Model tertutup dapat sulit direproduksi karena versi dan data pelatihan tidak terbuka. Model terbuka memberi kendali lebih tetapi memerlukan keahlian, komputasi, dan pemeliharaan. Tidak ada pilihan yang bebas biaya atau risiko. Peneliti memilih menurut data, sensitivitas, kebutuhan stabilitas, serta kemampuan tim. Ketidakpastian penyedia ditulis sebagai keterbatasan.

Jejak energi dan sumber daya komputasi menjadi pertimbangan terutama pada eksperimen berskala besar. Peneliti membedakan kebutuhan metodologis dari eksperimen berulang yang tidak menambah nilai. Model yang lebih besar tidak otomatis menghasilkan kesimpulan yang lebih baik. Penggunaan yang hemat dapat menurunkan biaya dan dampak lingkungan sambil meningkatkan fokus pada tujuan riset.

26. Agenda pengembangan metodologi

Penelitian berikutnya perlu menguji penggunaan AI pada berbagai tipe data, bahasa, institusi, serta disiplin dengan benchmark yang transparan. Studi komparatif harus melaporkan metode yang gagal, bukan hanya contoh terbaik. Penilaian perlu memisahkan akurasi, keterlacakan, waktu, biaya validasi, fairness, dan kegunaan keputusan. Dataset evaluasi harus mencakup kasus ambigu dan kelompok kurang terwakili.

Studi longitudinal menilai apakah penggunaan AI meningkatkan atau mengikis keterampilan analisis, bagaimana tugas tim berubah, dan siapa yang menanggung kerja pengecekan. Penelitian implementasi memeriksa apakah model benar-benar memperbaiki outcome pendidikan atau organisasi. Studi kausal dibutuhkan untuk menguji dampak intervensi AI, bukan hanya persepsi pengguna atau korelasi pemakaian.

Standar pelaporan dapat berkembang melalui kolaborasi lintas disiplin: nama model, versi, jenis tugas, data, preprocessing, prompt, output yang dipakai, validasi, keputusan manusia, dan batas berbagi. Panduan harus cukup fleksibel untuk analisis kualitatif maupun kuantitatif, tetapi cukup konkret untuk audit. Komunitas penelitian Indonesia memiliki peluang membangun benchmark bahasa dan data yang dikelola secara adil, dengan keterlibatan komunitas sumber data.

27. Memilih unit analisis dan menyiapkan struktur data

Pemilihan unit analisis merupakan keputusan yang terlihat sederhana tetapi menentukan makna seluruh tabel dan model. Satu baris data dapat mewakili seorang mahasiswa, satu mata kuliah, satu pendaftaran, satu semester, satu jawaban survei, atau satu interaksi digital. Jika data disusun pada tingkat peristiwa, seorang mahasiswa dapat muncul ratusan kali; jika model kemudian memperlakukan setiap baris sebagai individu independen, ketidakpastian akan dihitung secara keliru. Sebelum analisis, peneliti menggambar relasi entitas: siapa atau apa objek utama, bagaimana tabel terhubung, apakah hubungan satu-ke-satu atau satu-ke-banyak, dan variabel mana yang berubah dari waktu ke waktu.

Struktur data lebar menempatkan banyak variabel pada satu baris; struktur panjang menyimpan satu observasi variabel per baris. Keduanya sah untuk tujuan berbeda, tetapi transformasi dari satu bentuk ke bentuk lain dapat menggandakan atau menghilangkan observasi bila kunci tidak unik. AI dapat menulis kode reshape atau join, namun harus diberikan definisi kunci yang tepat. Pengguna perlu memeriksa jumlah baris sebelum dan sesudah penggabungan, jumlah kunci duplikat, dan proporsi kasus yang tidak berpasangan. Perbedaan ukuran bukan sekadar persoalan teknis: ia dapat mengubah denominator, komposisi sampel, dan bobot relatif peserta.

Data longitudinal menuntut identifikasi waktu dan urutan yang konsisten. Tanggal pendaftaran berbeda dari tanggal kejadian; timestamp server dapat memakai zona waktu lain; semester mungkin melintasi tahun kalender. Variabel yang tercatat setelah outcome tidak boleh digunakan seolah-olah tersedia saat prediksi dibuat. Peneliti membuat kamus data yang menerangkan nama, label, satuan, rentang yang mungkin, kategori khusus, sumber, waktu pencatatan, transformasi, dan pemilik data. Kamus ini menjadi penghubung antara kode, makna substantif, dan batas penggunaan.

Kualitas pengukuran juga perlu dinilai. Reliabilitas merujuk pada konsistensi pengukuran, sedangkan validitas menyangkut apakah skor memang merepresentasikan konstruk yang dituju. Nilai ujian dapat mengukur kinerja pada ujian tersebut, tetapi belum tentu kompetensi yang lebih luas. Jumlah login dapat mengukur aktivitas sistem, bukan perhatian atau pembelajaran. Model yang menemukan hubungan konsisten antara indikator dan outcome tidak mengatasi ketidakjelasan konstruk. Peneliti membandingkan indikator dengan teori, prosedur pengumpulan, hasil validasi terdahulu, serta perspektif peserta. Jika definisi berubah antar kohort, analisis perlu menyatakan apakah skor dapat dibandingkan.

Dalam data gabungan dari beberapa sistem, perbedaan skema dan kebijakan pengumpulan menjadi sumber bias. Satu fakultas mungkin mencatat kehadiran setiap pertemuan, yang lain hanya pada ujian; platform belajar dapat merekam klik secara berbeda; unit administrasi dapat mengubah definisi status aktif. Harmonisasi harus mempertahankan jejak asal. Peneliti menyimpan variabel sumber dan variabel harmonisasi secara terpisah, lalu menuliskan aturan pemetaan. Keputusan menyamakan dua kategori harus dilandasi kesetaraan makna, bukan sekadar kemiripan label. Sistem AI dapat menyarankan pemetaan istilah, tetapi tidak memiliki mandat untuk menyatakan dua praktik institusional ekuivalen tanpa verifikasi ahli.

28. Data hilang, pencilan, dan keputusan pembersihan

Data tidak lengkap adalah karakteristik proses pengukuran, bukan hanya gangguan yang perlu dibuang. Nilai dapat hilang karena peserta melewatkan pertanyaan, sensor gagal, variabel belum diberlakukan, atau suatu kelompok tidak tercakup sistem. Mekanisme kehilangan yang berbeda memiliki konsekuensi berbeda. Missing completely at random berarti kehilangan tidak berkaitan dengan variabel teramati maupun tidak teramati; missing at random berarti dapat dijelaskan melalui informasi teramati; missing not at random

menunjukkan kemungkinan hubungan dengan nilai yang hilang itu sendiri. Dalam praktik, asumsi ini tidak selalu dapat diuji penuh, sehingga peneliti perlu menjelaskan alasan substantif dan menjalankan analisis sensitivitas.

Menghapus semua baris yang mengandung nilai kosong dapat mengurangi ukuran sampel dan mengubah komposisi kelompok. Mengisi nilai dengan rerata dapat mengecilkan variasi dan merusak hubungan. Imputasi berganda dapat mempertahankan ketidakpastian lebih baik dalam kondisi tertentu, tetapi memerlukan model imputasi yang masuk akal serta penyertaan variabel yang relevan. AI dapat menjelaskan opsi atau menyusun kode, namun pilihan prosedur bergantung pada desain dan mekanisme kehilangan. Laporan menyebut jumlah dan persentase missing per variabel, pola bersama, keputusan, asumsi, serta apakah hasil berubah pada pendekatan lain. Bila data sensitif, ringkasan pola sebaiknya dibuat lokal atau dalam layanan institusi yang disetujui.

Pencilan juga tidak otomatis merupakan kesalahan. Nilai yang ekstrem dapat menandai salah input, unit yang berbeda, populasi khusus, atau fenomena penting yang memang ingin dipahami. Menyaring titik di luar rentang statistik tanpa memeriksa catatan asli berisiko menghapus subkelompok yang paling informatif. Pemeriksaan dilakukan melalui aturan domain, plot distribusi, jejak sumber, dan, bila diizinkan, verifikasi dengan pengelola data. Analisis utama dapat dilaporkan bersama sensitivitas yang memasukkan dan mengecualikan pencilan menurut kriteria yang ditetapkan. Pemangkasan atau winsorization tidak menyembuhkan salah spesifikasi; ia mengubah besaran yang sedang diringkas.

Kesalahan format sering tersembunyi di balik tabel yang tampak bersih. Nilai “0” dapat berarti nol nyata atau belum diisi. Tanggal dapat dibaca sebagai bulan-hari pada satu mesin dan hari-bulan pada mesin lain. Kategori “lainnya” dapat menyimpan teks bebas yang memuat identitas. Desimal koma dan titik

dapat dibaca berbeda, dan tanda minus dapat hilang dalam OCR. Skrip pembersihan sebaiknya mengandung pemeriksaan eksplisit, seperti jumlah nilai unik yang tidak diharapkan, minimum-maksimum, tipe variabel, dan distribusi sebelum-sesudah. AI dapat mengusulkan pemeriksaan, tetapi setiap aturan pembersihan harus dapat dijelaskan, ditinjau, dan dijalankan ulang.

29. Deskripsi, eksplorasi, dan pembentukan hipotesis

Analisis deskriptif menggambarkan apa yang tercatat dalam data. Ia tidak membutuhkan klaim kausal untuk berguna: komposisi responden, tren waktu, distribusi skor, tingkat respons, atau perbedaan jumlah layanan dapat membantu institusi memahami situasi. Namun statistik ringkas selalu memiliki denominator. Persentase peserta yang lulus bisa memakai semua mahasiswa terdaftar, hanya peserta ujian, atau hanya mahasiswa yang tidak mengundurkan diri. Grafik yang tidak menunjukkan denominator dan periode dapat menyesatkan meskipun perhitungannya tepat. Peneliti mencantumkan populasi, rentang waktu, definisi outcome, dan jumlah observasi.

Eksplorasi data membantu menemukan pola, kesalahan, kelompok, dan pertanyaan lanjutan. Visualisasi awal, korelasi, distribusi, atau teknik embedding dapat mengungkap nilai yang salah kode maupun struktur yang tak terduga. Akan tetapi, pola yang ditemukan sesudah mencoba banyak kombinasi bukan bukti konfirmatori dengan kekuatan yang sama seperti hipotesis yang direncanakan sebelum melihat data. Catatan analisis membedakan tahap eksploratif dari pengujian konfirmatori. Peneliti dapat membuat hipotesis baru dari eksplorasi lalu mengujinya menggunakan data lain, gelombang berikutnya, atau desain lanjutan.

Model generatif dapat bertindak sebagai mitra eksplorasi: menyarankan plot, mengidentifikasi kemungkinan segmentasi, atau menyusun pertanyaan diagnostik. Keluaran tersebut diperlakukan sebagai daftar kemungkinan, bukan temuan. Untuk setiap pola, peneliti kembali ke baris data, mengecek

unit, denominator, missingness, dan bentuk distribusi. Jika model mengatakan ada hubungan yang kuat, tim menghitung ukuran efek dan interval ketidakpastian secara independen. Jika model membuat narasi mengenai subkelompok, tim memeriksa apakah kategori itu benar-benar tercatat atau diimajinasikan dari konteks.

Pencarian pola tidak boleh berhenti pada satu angka seperti p-value. Ukuran efek dan interval ketidakpastian membantu pembaca memahami besaran serta presisi; visualisasi memperlihatkan variasi; analisis sensitivitas menguji ketahanan terhadap keputusan yang masuk akal. Uji statistik dapat menjawab pertanyaan probabilistik di bawah asumsi tertentu, tetapi tidak menilai apakah dampak penting secara pendidikan atau sosial. Perbedaan kecil dapat signifikan dengan sampel besar, sedangkan efek praktis dapat luput bila sampel kecil. Peneliti menyampaikan batas daya dan ketidakpastian, tidak mengubah “tidak signifikan” menjadi bukti bahwa tidak ada efek.

30. Menentukan metode inferensial dan menghormati desain

Metode inferensial dipilih karena struktur desain mendukung estimasi tertentu, bukan karena chatbot menyarankan uji yang populer. Data dari pengukuran berulang pada individu yang sama tidak setara dengan sampel independen; peserta yang dikelompokkan pada kelas atau fakultas dapat berbagi lingkungan; survei dengan bobot memerlukan analisis yang memperhitungkan rancangan sampel. Jika struktur kelompok diabaikan, standar error, interval, dan kesimpulan dapat menjadi terlalu optimistis. Peneliti menuliskan asumsi independensi, bentuk model, penanganan pengelompokan, kovariat, serta alasan substantif memasukkan atau tidak memasukkannya.

Perbandingan kelompok perlu memeriksa bagaimana kelompok terbentuk. Perbedaan nilai antara peserta program bimbingan dan nonpeserta dapat berasal dari seleksi mandiri: mahasiswa yang mengambil bimbingan mungkin

sejak awal lebih membutuhkan dukungan atau lebih termotivasi. Penyesuaian kovariat dapat mengurangi confounding yang terukur, tetapi tidak secara ajaib menghilangkan faktor tak terukur. Setiap klaim sebab-akibat harus didukung oleh desain, asumsi identifikasi, urutan waktu, dan pemeriksaan alternatif. Perangkat AI dapat mengajukan model regresi, tetapi peneliti bertanggung jawab menjelaskan makna koefisien dan estimand yang dihasilkan.

Multiplikitas juga muncul saat ribuan outcome, kelompok, atau spesifikasi dijelajahi. Jika pengujian berulang tidak diperhitungkan, sebagian hasil dapat tampak ekstrem secara kebetulan. Strateginya dapat berupa koreksi pengujian, hierarki hipotesis, analisis multilevel, validasi di sampel independen, atau fokus pada estimasi dan ketidakpastian alih-alih ambang biner. Pendekatan dipilih sesuai tujuan, bukan diterapkan mekanis. AI yang menghasilkan tabel puluhan p-value tanpa memperjelas bahwa banyak uji telah dicoba berisiko membuat pembaca memilih hasil yang paling menarik.

Kesesuaian perangkat lunak dan metode harus diverifikasi. Peneliti memeriksa dokumentasi fungsi, versi paket, pilihan default, penanganan nilai kosong, koreksi kontinuitas, dan cara menghitung interval. Kode yang dibuat model ditelaah sebagai kode yang ditulis oleh kolega: jalankan pada contoh kecil, periksa dimensi objek, uji kasus batas, bandingkan dengan hasil manual atau implementasi lain, dan pastikan output sesuai teori. Jika perangkat menyembunyikan keputusan otomatis, laporan menjelaskan opsi dan versi yang digunakan. Ketika perbedaan hasil muncul, bukan angka yang dipilih berdasarkan preferensi; penyebabnya ditelusuri dan dilaporkan.

31. Menilai model prediktif secara menyeluruh

Model prediktif harus dibandingkan dengan baseline yang bermakna. Untuk risiko kelulusan, baseline dapat berupa prevalensi rata-rata, skor awal, atau aturan yang saat ini digunakan. Model kompleks patut dipertahankan hanya jika peningkatan kinerjanya relevan terhadap keputusan dan biaya

operasional. Menyajikan akurasi saja tidak cukup. Jika 95 persen mahasiswa berhasil, model yang selalu memprediksi berhasil memperoleh akurasi 95 persen tetapi gagal mengidentifikasi mereka yang memerlukan dukungan. Precision, recall, specificity, area under curve, kalibrasi, dan manfaat keputusan menjawab aspek yang berbeda.

Ambang klasifikasi adalah pilihan kebijakan. Menurunkan threshold dapat menangkap lebih banyak mahasiswa yang akan gagal, tetapi menghasilkan lebih banyak false positive dan mungkin membuat konselor kewalahan. Menaikkan threshold mengurangi beban tindak lanjut tetapi melewatkan sebagian mahasiswa yang membutuhkan bantuan. Penetapan threshold mempertimbangkan kapasitas dukungan, konsekuensi label, biaya kedua jenis kesalahan, dan masukan pihak terdampak. Tidak ada nilai threshold yang "objektif" tanpa tujuan dan kapasitas intervensi yang dinyatakan.

Kalibrasi menilai apakah, misalnya, di antara prediksi risiko 0,30 sekitar 30 persen benar-benar mengalami outcome dalam data evaluasi. Model dapat meranking kasus dengan baik namun memberikan probabilitas yang terlalu tinggi atau rendah. Kalibrasi yang memadai pada keseluruhan populasi belum memastikan kesesuaian pada program kecil. Karena itu, peneliti memeriksa kinerja dan kalibrasi per kelompok dengan kehati-hatian atas ukuran sampel dan privasi. Jika estimasi kelompok sangat tidak pasti, angka tidak diperlakukan sebagai ranking pasti; pembaca diberi interval dan konteks.

Evaluasi juga memerlukan pemisahan data yang mencerminkan lingkungan penggunaan. Validasi acak mungkin tidak menggambarkan kohort mendatang jika kebijakan, kurikulum, atau komposisi mahasiswa berubah. Uji temporal pada tahun berikutnya lebih dekat dengan tujuan operasional. Validasi eksternal di institusi lain menguji transportabilitas, tetapi memerlukan harmonisasi konstruk dan definisi. Dataset uji tidak digunakan berulang kali untuk mengarahkan semua keputusan model karena secara bertahap ia

menjadi bagian dari proses pengembangan. Bila dataset terbatas, validasi silang dilakukan dengan prosedur yang menjaga struktur dan ketergantungan.

Model yang dikembangkan bukan model yang selesai. Setelah diterapkan, perubahan distribusi input, outcome, layanan, dan perilaku pengguna dapat mengubah kinerja. Intervensi yang berhasil bahkan dapat mengubah pola data: mahasiswa menerima bimbingan sehingga risiko aktual menurun. Pemantauan harus membedakan drift data dari perubahan manfaat program. Sistem menetapkan jadwal tinjauan, kriteria penghentian, pemilik yang bertanggung jawab, pencatatan insiden, serta prosedur untuk memperbarui atau menarik model. Skor harus memiliki tanggal kedaluwarsa dan konteks penggunaan, bukan menjadi atribut tetap seseorang.

32. Validasi analisis teks dan hasil kualitatif

Penggunaan AI untuk membaca teks akademik atau wawancara membawa pertanyaan interpretatif. Respons seseorang dapat memiliki ironi, ambiguitas, dialek, istilah lokal, atau makna yang bergantung pada percakapan sebelumnya. Model cenderung menghasilkan kategori yang koheren dan meyakinkan, tetapi kerapian label tidak membuktikan kecocokan dengan pengalaman peserta. Peneliti mulai dari pertanyaan dan kerangka epistemologis: apakah tujuan analisis mengidentifikasi frekuensi tema, memetakan variasi, memahami pengalaman hidup, atau membangun teori proses? Masing-masing mengizinkan jenis kompresi dan automasi yang berbeda.

Jika kodebook telah ditetapkan, LLM dapat melakukan klasifikasi awal atau membantu menemukan contoh kandidat. Sampel dikode ulang secara independen oleh manusia, termasuk kasus negatif dan ambigu. Tim menghitung ketepatan per kategori bila label acuan dapat didefinisikan, tetapi tidak mereduksi seluruh kualitas interpretasi menjadi satu metrik agreement. Pemeriksaan juga melihat kecenderungan model mengabaikan konteks,

menggabungkan tema yang terpisah, atau memberi label berdasarkan stereotip. Kutipan sumber harus ditampilkan sehingga pembaca dapat menilai hubungan antara data dan interpretasi.

Dalam analisis induktif, penggunaan model sebagai pembentuk tema dapat mengarahkan perhatian ke struktur yang sudah dikenal model dan menjauh dari kategori yang khas bagi konteks lokal. Peneliti merekam bagaimana tema muncul, bagaimana alternatif ditolak, dan bagaimana posisi peneliti memengaruhi interpretasi. AI dapat menawarkan pembacaan tandingan atau mencari contoh yang tidak cocok, tetapi ia tidak mempunyai pengalaman sosial peserta maupun refleksivitas metodologis peneliti. Ia tidak boleh mengubah ucapan yang ragu-ragu menjadi pernyataan tegas demi ringkasan yang lancar.

Pertimbangan privasi teks sering lebih sulit dari sekadar menghapus nama. Kombinasi jabatan, lokasi, peristiwa, dan ungkapan unik dapat mengidentifikasi pembicara. Mengunggah kutipan ke sistem publik tetap berpotensi memindahkan informasi meski identitas langsung disamarkan. Peneliti menilai layanan pemrosesan, retensi, akses, dan izin protokol. Jika tidak ada jalur yang memadai, analisis dilakukan lokal atau AI tidak digunakan pada materi tersebut. Model juga dapat mengutip ulang informasi sensitif dalam narasi keluarannya, sehingga hasil disunting sebelum dibagikan.

Kesimpulan kualitatif yang dapat dipercaya menunjukkan jejak analitis: potongan data, kode, pengelompokan, tema, kasus penyangkal, serta interpretasi yang menghubungkan data ke pertanyaan dan teori. Ringkasan otomatis berguna untuk orientasi, tetapi bentuk singkatnya dapat menghapus perbedaan suara. Karena itu, naskah menyampaikan variasi dan ketegangan, bukan hanya daftar tema yang tampak bersih. Temuan Jalali dan Akhavan (2024) serta Nguyen-Trung (2025) menjadi pengingat bahwa bantuan LLM dalam analisis kualitatif perlu dibatasi dan dikaji pada konteks tugas tertentu.

33. Menyajikan visualisasi dan narasi kuantitatif

Visualisasi merupakan bagian dari argumentasi. Pemilihan sumbu, skala, warna, urutan kategori, rentang waktu, dan denominator menentukan apa yang mudah dilihat. Grafik batang yang dimulai dari nilai bukan nol dapat memperbesar perbedaan; garis tren yang memotong periode penting dapat menyiratkan kesinambungan yang tidak ada; warna merah dan hijau dapat menyulitkan pembaca dengan gangguan persepsi warna. AI dapat membuat kode grafik dengan cepat, tetapi peneliti memeriksa apakah gambar menyampaikan data secara jujur dan mudah diakses.

Grafik harus membantu menjawab pertanyaan tertentu. Distribusi lebih baik ditampilkan dengan plot yang memperlihatkan variasi daripada hanya rerata. Perbandingan kelompok dapat memperlihatkan ukuran sampel dan interval ketidakpastian. Data longitudinal sebaiknya mencantumkan waktu dan cohort dengan jelas. Peta harus menjelaskan satuan ruang dan metode agregasi. Jika data sedikit atau kategori sangat kecil, tampilan visual dapat membuka identitas, sehingga privasi menjadi pertimbangan desain.

Narasi hasil berisi tiga lapisan yang berbeda: apa yang dihitung, apa arti ukuran tersebut dalam konteks substantif, dan apa yang dapat disimpulkan menurut desain. Misalnya “rerata skor meningkat 4 poin” adalah deskripsi; “peningkatan tersebut mungkin bermakna dalam konteks mata kuliah” adalah interpretasi; “program menyebabkan kenaikan” adalah klaim kausal yang memerlukan desain pendukung. Model generatif sering merangkai ketiganya menjadi cerita lancar. Peneliti menguraikan kembali kalimat model dan menghapus frasa yang melebihi bukti.

Laporan sebaiknya memuat hasil nol atau campuran yang relevan. Pemilihan grafik dan teks setelah melihat semua keluaran perlu dijelaskan bila dapat mengubah penekanan. Hasil utama ditautkan ke tabel atau kode sumber, sementara informasi tambahan disediakan dengan cara yang aman. Pembaca

tidak seharusnya harus mempercayai kalimat ringkasan AI tanpa dapat melihat definisi, ukuran, atau batas analisis. Aksesibilitas dan transparansi memperluas siapa yang dapat mengkritik temuan.

34. Dokumentasi, provenance, dan reproduksibilitas dengan AI

Reproduksibilitas membutuhkan pencatatan bukan hanya kode akhir, tetapi juga asal data, versi variabel, transformasi, parameter, perangkat lunak, random seed bila relevan, dan lingkungan eksekusi. AI menambah beberapa detail: nama penyedia dan model, tanggal atau versi layanan, pengaturan sampling, prompt substantif, dokumen yang diberikan, output yang dipakai, serta proses penyuntingan. Detail prompt dapat memuat rahasia atau data sensitif; arsip harus mengikuti pengaturan akses dan minimisasi informasi.

Simpan data mentah secara read-only dan pisahkan dengan jelas dari data bersih. Setiap tahap menghasilkan artefak baru dan log yang menerangkan jumlah baris, aturan transformasi, serta peringatan. Data pribadi tidak harus dipublikasikan demi reproduksibilitas; peneliti dapat menyediakan kode, data sintetis, protokol akses terkendali, atau statistik agregat. Prinsip FAIR mendorong data dapat ditemukan, diakses dengan aturan yang sesuai, interoperabel, dan dapat digunakan kembali, tetapi tidak mengesampingkan privasi dan persetujuan.

Jika model bahasa membantu menulis skrip, skrip yang telah diverifikasi menjadi artefak analisis. Menyimpan percakapan saja tidak cukup karena prompt yang sama dapat memunculkan output berbeda, model dapat diperbarui, dan kode dapat mengandung kesalahan. Reproduksi yang kuat menyimpan kode, hasil pemeriksaan, data versi yang diizinkan, serta dokumentasi dependensi. Untuk pemanggilan model yang menjadi bagian transformasi inti, catat juga output mentah yang dipakai dan hash atau identifier yang mendukung pemeriksaan, jika kebijakan memungkinkan.

Provenance menjawab pertanyaan “bagaimana angka ini terbentuk?”. Suatu hasil sebaiknya dapat ditelusuri dari paragraf laporan ke tabel atau visual, dari tabel ke objek data dan skrip, serta dari skrip ke sumber. Pelacakan ini membantu menemukan kapan definisi atau aturan berubah. Peninjau internal dapat mengulangi sebagian operasi pada sampel untuk menguji apakah instruksi cukup jelas. Ketika data tidak dapat dibuka, dokumentasi tertutup yang disediakan melalui proses aman masih dapat memperkuat audit.

Reproduksibilitas juga sosial dan organisasi. Keahlian untuk menjalankan kode, hak akses, stabilitas layanan, serta kontinuitas staf menentukan apakah hasil dapat dipertahankan. Tim menulis dokumentasi untuk pembaca yang bukan pembuat skrip, menggunakan bahasa yang menjelaskan tujuan tiap transformasi, dan menghindari ketergantungan pada pengetahuan tacit seorang analis. AI dapat menyusun dokumentasi awal, tetapi pemilik pipeline memeriksa akurasinya. Pedoman evaluasi machine learning menempatkan leakage dan prosedur eksperimental sebagai masalah inti, sehingga kualitas pelacakan sama pentingnya dengan pemilihan algoritma (Kapoor & Narayanan, 2023).

35. Menghubungkan analisis dengan keputusan dan dampak

Analisis sering dipakai untuk menyarankan keputusan; transisi itu perlu dirancang secara eksplisit. Model risiko mahasiswa mungkin ditujukan untuk memberi prioritas pada konseling, bukan menentukan kelayakan beasiswa. Jika hasil yang sama digunakan untuk tujuan berbeda tanpa validasi baru, batas model dilanggar. Dokumen penggunaan menjelaskan pengguna yang diizinkan, waktu prediksi, variabel yang diperlukan, populasi sasaran, tindakan yang diperbolehkan, dan hal yang dilarang.

Sebelum implementasi, tim menilai apakah prediksi dapat diubah menjadi bantuan nyata. Peringatan tanpa kapasitas layanan dapat menambah kecemasan dan menghasilkan beban yang tak tertangani. Label yang terlihat

oleh banyak pihak dapat memengaruhi harapan dosen terhadap mahasiswa. Pelaksanaan yang tepat menyediakan jalur kontak yang suportif, kesempatan mengoreksi data, informasi yang mudah dipahami, serta pilihan untuk tidak menggunakan output otomatis sebagai satu-satunya dasar tindakan. Dampak ini adalah bagian dari analisis karena menentukan apakah penggunaan sistem bermanfaat atau merugikan.

Evaluasi pascaimplementasi mengukur konsekuensi pada pengguna dan outcome, bukan hanya throughput. Apakah intervensi menjangkau kelompok yang sebelumnya kurang terlayani? Apakah dukungan meningkatkan keberhasilan, atau hanya meningkatkan jumlah flag? Apakah pengguna mengerti skor? Apakah model menciptakan perubahan perilaku yang mengurangi keterbukaan mahasiswa? Wawancara, audit keputusan, metrik, dan kajian insiden dapat digabungkan. Perubahan protokol dilakukan bila ditemukan ketidakadilan atau penyalahgunaan.

Keterlibatan komunitas menambah kualitas interpretasi. Mahasiswa, dosen, staf, atau peserta survei mengetahui konsekuensi bahasa dan indikator yang dapat luput dari tim teknis. Konsultasi bukan sekadar persetujuan simbolis; masukan dapat mengubah definisi outcome, dashboard, akses, atau keputusan untuk tidak melakukan prediksi individual. Jika model memengaruhi kelompok yang sejak awal rentan, mereka perlu dilibatkan sejak tahap tujuan dan desain. Tata kelola yang adil mengakui bahwa akurasi teknis bukan satu-satunya dasar legitimasi.

Analisis yang tidak diikuti tindakan juga memiliki nilai jika memperjelas pengetahuan, menolak dugaan yang salah, atau menunjukkan data yang belum cukup. Peneliti sebaiknya tidak mencari intervensi algoritmik untuk setiap pola. Kadang hasil paling bertanggung jawab adalah memperbaiki pencatatan, mengubah desain layanan universal, mengumpulkan data

kualitatif, atau tidak memakai data untuk klasifikasi. AI mempercepat pemrosesan, namun pilihan respons tetap keputusan manusia dan institusi.

36. Ukuran sampel, daya, dan ketidakpastian

Ukuran sampel yang memadai tidak dapat ditentukan hanya dengan aturan umum seperti “sepuluh observasi per variabel”. Kebutuhan bergantung pada pertanyaan, rancangan, variasi outcome, ukuran efek yang secara substantif penting, struktur pengelompokan, jumlah parameter, prevalensi kelas, dan kesalahan yang ditoleransi. Untuk studi uji hipotesis, perencanaan dapat dimulai dari efek minimum yang ingin dideteksi, taraf kesalahan, daya yang diinginkan, serta model analisis. Untuk model prediktif, pertimbangan mencakup jumlah outcome positif, kompleksitas model, stabilitas estimasi, dan ukuran data yang diperlukan untuk menguji performa secara independen. AI dapat menghitung skenario jika asumsi diberikan, tetapi tidak mengetahui efek mana yang layak secara kebijakan atau akademik.

Perencanaan yang jujur menyatakan sumber asumsi: studi terdahulu, pilot, pertimbangan substantif, atau skenario konservatif. Jika nilai parameter tidak diketahui, tim menghitung beberapa skenario dan memperlihatkan bagaimana kebutuhan sampel berubah. Peneliti tidak memilih asumsi yang memberi angka paling mudah dicapai. Pada penelitian pendidikan, jumlah baris sangat berbeda dari jumlah unit independen: seribu rekaman dari dua puluh kelas tetap memiliki ketergantungan dalam kelas. Rancangan berkelompok memerlukan perhatian terhadap intraclass correlation, jumlah cluster, dan ukuran tiap cluster. Memecah satu kelompok kecil ke dalam train/test dapat membuat set uji tidak informatif.

Pada studi kualitatif, logika ukuran sampel berbeda. Kecukupan data berhubungan dengan tujuan, variasi peserta, kualitas wawancara, strategi sampling, dan kedalaman interpretasi. Sistem generatif yang merangkum banyak wawancara tidak menciptakan pengalaman tambahan dan tidak

mengubah sampel purposif menjadi representatif statistik. Model dapat memberi kesan “saturasi” setelah mengelompokkan tema, tetapi keputusan mengenai kedalaman konsep memerlukan argumentasi metodologis. Peneliti menerangkan bagaimana partisipan dipilih, informasi yang dicari, dan alasan data dianggap cukup untuk klaim yang dibuat.

Ketidakpastian hadir pada beberapa lapisan: variasi sampling, pengukuran tidak sempurna, pemilihan model, pembersihan data, asumsi, dan ketidakstabilan output generatif. Interval statistik menangkap sebagian ketidakpastian, tetapi tidak seluruhnya. Analisis sensitivitas dapat menilai perubahan hasil saat definisi outcome, penanganan missing, kovariat, atau kelompok analisis diganti secara masuk akal. Jika beberapa pilihan menghasilkan kesimpulan berbeda, perbedaan tersebut dilaporkan, tidak disembunyikan dengan memilih spesifikasi paling menyenangkan. Dalam analisis kualitatif, ketidakpastian dijelaskan melalui kasus yang ambigu, perbedaan interpretasi, dan batas cakupan.

AI sering menulis nilai pasti dengan gaya yakin, padahal input atau metode mungkin tidak cukup. Peneliti memisahkan presisi numerik dari kepastian ilmiah. Nilai sampai enam desimal tidak membuat estimasi lebih akurat dari mutu datanya. Bahasa probabilistik dijaga secara konsisten: “terkait”, “diprediksi”, atau “konsisten dengan” dipilih sesuai desain, sedangkan “menyebabkan” memerlukan dasar kausal. Pelaporan ketidakpastian membantu pembaca membuat keputusan proporsional dan mencegah angka prediksi diperlakukan sebagai fakta individu.

37. Bias, fairness, dan perbedaan dampak antar kelompok

Bias algoritmik dapat muncul di setiap titik rantai penelitian. Cakupan data mungkin tidak seimbang; ukuran dapat lebih akurat untuk sebagian kelompok; label historis dapat memuat ketidakadilan; variabel proksi dapat merepresentasikan sumber daya; dan penerapan model dapat mengubah

perilaku orang yang diprediksi. Karena itu, "menghapus atribut sensitif" tidak otomatis membuat sistem adil. Kode pos, pola interaksi, institusi asal, atau perangkat yang digunakan dapat bertindak sebagai proksi. Pengujian harus memeriksa baik atribut yang digunakan maupun mekanisme sosial yang menghasilkan fitur dan outcome.

Konsep fairness tidak selalu sejalan. Kesetaraan tingkat positif, keseimbangan error, kalibrasi, dan akses manfaat adalah sasaran yang berbeda dan dalam kondisi tertentu sulit dipenuhi bersama, terutama ketika prevalensi outcome berbeda antar kelompok. Peneliti menghindari klaim abstrak bahwa suatu model "fair" tanpa menyebut definisi, populasi, threshold, dan konteks keputusan. Metrik dapat menandai kesenjangan, tetapi penilaian normatif tentang tradeoff melibatkan pihak terdampak dan institusi yang bertanggung jawab. Perbedaan metrik bukan sekadar masalah statistik; pilihan tersebut menentukan siapa yang menerima false positive atau false negative.

Audit fairness dimulai dari pemetaan jalur keputusan. Siapa yang tercakup, siapa yang tidak terwakili, siapa mengumpulkan data, kategori apa yang digunakan, dan apa tindakan setelah skor keluar? Kemudian tim memeriksa kualitas sampel, missingness, kesalahan pengukuran, distribusi outcome, performa per kelompok, kalibrasi, dan konsekuensi threshold. Bila ukuran kelompok kecil, estimasi metrik dapat sangat tidak stabil dan risiko pengungkapan meningkat. Solusinya bukan menghapus kelompok itu dari kajian, melainkan menyatakan keterbatasan, menggunakan perlindungan statistik yang sesuai, melibatkan komunitas, dan tidak membuat keputusan individual berisiko tinggi berdasarkan bukti rapuh.

Keadilan juga bergantung pada tujuan. Model yang lebih baik dalam mendeteksi kebutuhan bantuan belum tentu adil jika hanya memberi dukungan kepada mahasiswa yang sudah sering berinteraksi dengan sistem digital. Model yang memprediksi keluar kuliah dapat memperkuat stigma jika

label dibagikan tanpa penjelasan dan dukungan. Analisis mempertimbangkan apakah keputusan memperluas sumber daya atau mengurangi kesempatan. Pemantauan efek umpan balik penting: kelompok yang lebih sering ditandai dapat lebih diawasi, menghasilkan lebih banyak catatan negatif, lalu tampak semakin berisiko.

Contoh pada layanan kesehatan memperlihatkan bahwa outcome proksi yang tidak tepat dapat menghasilkan bias sistematis meskipun algoritma mengikuti prosedur komputasi (Obermeyer et al., 2019). Pelajaran metodologisnya berlaku luas: model mengoptimalkan target yang dirumuskan, bukan nilai yang sebenarnya diinginkan masyarakat. Tim menilai kesesuaian target, membuka ruang keberatan, dan menetapkan penghentian saat manfaat tidak seimbang. Hasil fairness disajikan bersama ketidakpastian dan interpretasi substantif, bukan sebagai skor sertifikasi tunggal.

38. Analisis data multimodal dan integrasi sumber

Penelitian kini dapat menggabungkan teks, tabel, audio, gambar, video, sensor, dan catatan administratif. Model multimodal menawarkan kemampuan menghubungkan kanal tersebut, tetapi setiap sumber memiliki unit, resolusi, cakupan waktu, serta error pengukuran yang berbeda. Foto ruang kelas tidak langsung mengukur keterlibatan; rekaman suara tidak otomatis menangkap pemahaman; sensor aktivitas mencatat sinyal fisiologis atau gerak dalam konteks tertentu, bukan makna pengalaman. Peneliti mendefinisikan konstruk, persetujuan, dan hubungan bukti sebelum menyatukan sumber.

Sinkronisasi waktu dapat menjadi tantangan utama. Timestamp kamera, LMS, dan survei mungkin berbeda zona waktu atau granularitas. Sejumlah sensor mengirim data terus-menerus, sedangkan outcome diukur hanya setiap semester. Menggabungkan agregat mingguan dengan nilai akhir harus mempertimbangkan periode yang relevan dan kemungkinan kebocoran informasi setelah outcome. Pipeline menyimpan aturan penyelarasan dan

tingkat agregasi. Jika data individu terhubung lintas kanal, kunci pencocokan dikelola dengan kontrol akses terpisah dan tujuan penggunaan yang jelas.

Representasi multimodal dapat memperkuat analisis sekaligus memperluas permukaan privasi. Suara atau wajah mungkin dapat mengidentifikasi seseorang walau nama dihapus; latar belakang video dapat mengungkap lokasi, keluarga, atau kondisi rumah. Data tersebut dapat mengandung sinyal sensitif yang tidak disadari peserta ketika menyetujui penelitian. Tim menjelaskan jenis inferensi yang mungkin dilakukan, mengurangi detail yang tidak dibutuhkan, membatasi masa simpan, dan menyediakan alternatif partisipasi bila memungkinkan. Model pihak ketiga tidak menerima materi sebelum persetujuan etik dan perlindungan data mencakup pemrosesan tersebut.

Evaluasi model multimodal memerlukan ablation atau pengujian kontribusi tiap kanal. Jika model menggabungkan teks dengan klik, tim memeriksa apakah kanal tambahan benar-benar meningkatkan kualitas prediksi di luar baseline dan untuk siapa. Missing modality dapat terkonsentrasi pada peserta dengan konektivitas rendah, sehingga sistem yang mengharuskan video bisa secara sistematis mengurangi keterwakilan. Peneliti membandingkan penggunaan semua kanal dengan pendekatan yang mengakomodasi data tidak lengkap, serta melaporkan efek perubahan komposisi.

Dalam metode campuran, integrasi bukan sekadar menumpuk dataset kuantitatif dan kutipan kualitatif. Joint display dapat mempertemukan tren, penjelasan pengalaman, kasus yang bertentangan, dan implikasi untuk teori. AI dapat mengusulkan penghubung atau ringkasan lintas sumber, tetapi peneliti memeriksa apakah sumber benar-benar merujuk pada unit dan fenomena yang sama. Ketidaksesuaian antara survei dan wawancara bisa menunjukkan perbedaan populasi, periode, atau definisi, bukan kesalahan

yang harus diselaraskan. Kontradiksi dilaporkan sebagai temuan yang berpotensi memperkaya pemahaman.

39. Menguji AI sebagai analis: benchmark dan evaluasi tugas

Evaluasi model perlu dibangun di sekitar tugas akademik yang spesifik. "Apakah AI mampu menganalisis data?" terlalu luas untuk satu angka kinerja. Tugas perlu menjelaskan bentuk input, kualitas dokumentasi, tingkat kesulitan, perangkat yang tersedia, bentuk output, kriteria keberhasilan, dan konsekuensi kesalahan. Menulis kode deskriptif dari tabel bersih berbeda dari memilih rancangan kausal, menjelaskan missing not at random, atau menafsirkan hasil kualitatif yang ambigu. Benchmark yang menggabungkan semua tugas menjadi skor agregat menyamarkan profil kemampuan dan keterbatasan.

Pembanding manusia juga perlu didefinisikan. Pengguna awam, mahasiswa metodologi, analis profesional, dan peneliti bidang dapat menghasilkan tingkat kinerja berbeda. Apakah waktu prompt dan revisi dihitung? Apakah manusia mendapat akses dokumentasi? Apakah jawaban dinilai oleh satu pakar atau beberapa penilai independen? Apakah kegagalan kode dihitung sebagai kesalahan atau dibiarkan pengguna memperbaikinya? Pelaporan eksperimen mencatat kondisi tersebut agar klaim tidak melampaui konteks. Studi eksploratif seperti Prandner et al. (2025) berguna justru ketika batas tugas dan kondisi penggunaan dijelaskan.

Kriteria evaluasi sebaiknya meliputi ketepatan numerik, kesesuaian metode, validitas inferensi, keterlacakan sumber, ketahanan terhadap perubahan prompt, kemampuan menyatakan ketidakpastian, dan beban verifikasi manusia. Pada ekstraksi data, perlu dihitung field yang tepat dan kesalahan kritis, bukan hanya nilai rerata. Pada coding kualitatif, kategori yang benar per kutipan tidak menangkap keseluruhan kualitas interpretasi; kutipan

penyangkal dan konsistensi definisi juga diperiksa. Pada analisis statistik, kode yang berhasil dijalankan belum tentu menjawab pertanyaan dengan benar.

Benchmark harus mengandung kasus batas yang realistis: label yang tidak jelas, variabel dengan kode khusus, kelompok kecil, data tidak seimbang, outcome yang ambigu, dan instruksi yang kurang lengkap. Sistem yang selalu memberi jawaban pasti dapat terlihat baik pada soal sederhana tetapi berbahaya ketika seharusnya meminta klarifikasi. Kemampuan abstain atau menyatakan "informasi tidak cukup" menjadi aspek kualitas. Pengujian juga meliputi robustness terhadap variasi ejaan, bahasa lokal, format file, dan perubahan istilah.

Perubahan model mengharuskan evaluasi berulang. Versi baru, retrieval berbeda, plugin kode, atau pengaturan sampling dapat mengubah perilaku. Tim menyimpan suite evaluasi internal yang tidak berisi data rahasia, menjalankan ulang uji sebelum pemakaian penting, dan mencatat regresi. Benchmark tidak menggantikan validasi atas proyek sesungguhnya; ia menunjukkan kapasitas umum pada contoh terdefinisi. Dalam artikel akademik, penulis menyebut kapan evaluasi berlangsung dan tidak mengasumsikan performa tahun ini berlaku selamanya.

40. Kedaulatan data, bahasa, dan konteks lokal

Analisis akademik di Indonesia berhadapan dengan ragam bahasa, campur kode, istilah institusi, praktik pengarsipan, dan kategori sosial yang tidak selalu diwakili data pelatihan global. Model dapat menerjemahkan istilah lokal ke kategori umum yang tidak tepat atau menghaluskan cara tutur sehingga perbedaan makna hilang. Prompt berbahasa Indonesia mungkin memberi output yang berbeda dari prompt berbahasa Inggris. Tim menguji padanan bahasa pada contoh representatif dan melibatkan pembaca yang memahami konteks, bukan menganggap keluaran bahasa Inggris sebagai standar kebenaran.

Kategori demografis harus dibentuk dengan sensitivitas terhadap sejarah dan tujuan penelitian. Kategori resmi dapat tidak sesuai dengan identitas yang digunakan komunitas, dan kategori yang terlalu rinci dapat memudahkan identifikasi. Peneliti menjelaskan mengapa pengelompokan tertentu diperlukan, bagaimana peserta mendeskripsikan diri, bagaimana nilai yang tidak cocok diperlakukan, dan apa keterbatasan agregasi. Sistem generatif tidak diberi kewenangan untuk menebak etnisitas, agama, disabilitas, atau identitas lain dari teks maupun nama kecuali desain penelitian, dasar etik, dan justifikasinya sangat kuat.

Kedaulatan data mencakup siapa yang menentukan tujuan pemrosesan, mengendalikan penyimpanan, mendapat akses, dan memperoleh manfaat dari temuan. Data lembaga atau komunitas tidak selalu menjadi bebas digunakan karena dapat diunduh. Perjanjian akses, persetujuan, kebijakan institusi, dan ekspektasi peserta tetap berlaku. Saat memakai layanan komersial, organisasi memahami apakah data disimpan, siapa subpemroses, apakah isi dapat digunakan untuk pelatihan, serta bagaimana penghapusan berlangsung. Untuk data berisiko, opsi lokal, pseudonimisasi, komputasi aman, atau tidak menggunakan model dapat lebih sesuai.

Ketimpangan infrastruktur juga perlu dipertimbangkan. Institusi dengan komputasi dan langganan mahal dapat memproses lebih banyak data dan menghasilkan publikasi lebih cepat, sementara universitas kecil dan komunitas lokal menanggung risiko ekstraksi data tanpa memperoleh manfaat. Replikasi dan benchmark terbuka dapat mengurangi ketimpangan bila tidak memaksa publikasi informasi sensitif atau kerja tanpa kompensasi. Kolaborasi menetapkan atribusi, kepemilikan, kapasitas lokal, dan akses terhadap alat serta hasil.

Konteks lokal memengaruhi interpretasi sebab, hubungan, dan dampak. Pola ketidakhadiran mungkin mencerminkan tanggung jawab kerja atau

transportasi, bukan rendahnya motivasi. Peringkat performa organisasi mungkin mengikuti kesempatan yang tidak merata, bukan kemampuan individu. AI dapat menyarankan penjelasan generik yang terdengar masuk akal, tetapi penjelasan substantif diperiksa terhadap sejarah, teori, dan suara pihak terdampak. Validitas bukan hanya apakah model bekerja di dataset; ia mencakup apakah kategori, tujuan, dan penggunaan bermakna bagi komunitas tempat penelitian dilakukan.

41. Contoh alur analisis dari pertanyaan sampai pelaporan

Bayangkan sebuah tim ingin mengetahui apakah pola penggunaan LMS pada semester pertama berkaitan dengan keberhasilan mata kuliah pengantar. Rumusan awal “apakah AI dapat memprediksi keberhasilan?” terlalu umum. Tim memperjelas populasi sebagai mahasiswa yang mengambil mata kuliah tersebut pada beberapa tahun akademik, outcome sebagai status lulus menurut catatan resmi, unit analisis sebagai mahasiswa per semester, serta waktu prediksi sebagai akhir minggu keempat. Definisi waktu mencegah penggunaan nilai akhir atau aktivitas setelah minggu keempat sebagai fitur. Tim juga membedakan tujuan prediksi untuk mengarahkan dukungan dari tujuan evaluasi sebab-akibat program.

Sebelum mengunduh data, peneliti memastikan persetujuan etik dan dasar akses, membatasi variabel pada yang dibutuhkan, serta menentukan siapa yang dapat melihat identitas. Kamus data mencatat apa arti login, aktivitas materi, pengumpulan tugas, dan status enrollment. Pengelola sistem menerangkan perubahan platform antar tahun. Data mentah disalin ke area read-only dengan identifier samaran; kunci identitas disimpan secara terpisah. Prosedur ini tidak hanya memenuhi tata kelola, melainkan juga memungkinkan tim menafsirkan perubahan distribusi ketika sistem kuliah berganti.

Audit eksploratif menemukan sebagian mahasiswa tercatat tanpa aktivitas digital, beberapa nilai waktu berada di luar periode semester, dan satu tahun memiliki definisi tugas yang berubah. Tim memeriksa apakah aktivitas nol merupakan tidak menggunakan LMS, akses melalui sistem lain, atau kegagalan pencatatan. Nilai waktu yang ganjil ditelusuri ke zona waktu server. Perubahan definisi tugas dipertahankan sebagai penanda tahun dan dibahas dalam validasi. Model generatif membantu menyusun kode pemeriksaan dan daftar plot, tetapi tim menjalankan audit dengan skrip yang direview, mencatat temuan, dan menyetujui aturan pembersihan sebelum memodelkan.

Kemudian tim membuat baseline yang mudah ditafsirkan, misalnya model sederhana menggunakan nilai diagnostik awal dan kehadiran yang memang tersedia pada minggu keempat. Model pembanding dapat berupa regresi logistik teratur dan metode pohon. Pemisahan data didasarkan pada cohort: tahun terdahulu untuk pengembangan, cohort berikutnya untuk validasi temporal. Pembagian acak setiap baris berisiko membuat pola semester yang sama muncul di train dan test. Tim melaporkan prevalensi lulus, ukuran cohort, distribusi fitur, dan batas generalisasi ke mata kuliah lain.

Hasil menunjukkan model pohon sedikit meningkatkan recall dibanding baseline, tetapi membutuhkan lebih banyak pemeriksaan dan performa turun pada satu program studi. Analisis kalibrasi menunjukkan probabilitas risiko terlalu tinggi pada sebagian mahasiswa. Alih-alih menerbitkan skor otomatis, tim melakukan pilot dukungan sukarela dengan threshold yang dirancang bersama konselor dan perwakilan mahasiswa. Dosen hanya mendapat daftar undangan yang membutuhkan tindak lanjut, bukan label permanen; mahasiswa diberi cara meminta koreksi. Evaluasi pilot mengukur kontak yang berhasil, akses layanan, outcome semester, dan efek pada kelompok.

Laporan akhir menyatakan bahwa aktivitas LMS memprediksi outcome dalam kondisi tertentu, tidak membuktikan bahwa menambah klik akan

menyebabkan kelulusan. Tim menyediakan kode pembersihan dan analisis tanpa identitas, menjelaskan variabel yang tidak tersedia, mencantumkan tanggal versi paket, serta menyimpan tabel agregat yang dapat direproduksi. Prompt yang membantu membuat kode disebutkan jika relevan, namun angka dihitung oleh skrip yang telah diuji. Contoh ini menunjukkan urutan: definisikan keputusan, audit asal data, tetapkan batas waktu informasi, bandingkan baseline, validasi secara temporal, evaluasi konsekuensi, lalu publikasikan klaim sesuai desain.

42. Menangani kegagalan dan membuat jejak audit

Dalam analisis berbantuan AI, kegagalan bukan peristiwa langka yang bisa diabaikan. Kode dapat salah menghitung kategori, output dapat mengarang nama kolom, model dapat meringkas tabel dengan denominator keliru, atau layanan dapat berubah sehingga jawaban tidak lagi cocok. Tim membuat kategori insiden mulai dari kesalahan ringan yang tertangkap sebelum dipakai hingga kesalahan yang memengaruhi keputusan atau publikasi. Tujuannya bukan mempermalukan individu, melainkan memastikan kesalahan terdeteksi, dampaknya diketahui, dan prosedur diperbaiki.

Kontrol teknis dapat mencakup pemeriksaan tipe dan rentang, jumlah baris, jumlah nilai unik, validasi skema, unit test untuk fungsi pembersihan, serta perbandingan hasil dengan contoh yang diketahui. Untuk narasi berbasis output, angka dan klaim utama ditautkan ke tabel sumber. Pemeriksa independen menelusuri sampel kasus dari catatan asli ke variabel turunan dan hasil akhir. Ambang pemeriksaan dibuat sebanding dengan risiko: koreksi ejaan pada draf tidak memerlukan proses sama seperti sistem yang memengaruhi akses layanan atau keputusan akademik.

Jika masalah ditemukan setelah publikasi, tim mengidentifikasi cakupan: bagian mana yang menggunakan output keliru, analisis mana yang terpengaruh, apakah kesimpulan berubah, serta siapa yang perlu diberi tahu.

Koreksi tabel mungkin cukup jika kesimpulan stabil; bila interpretasi utama berubah, jurnal, institusi, dan pihak terdampak menerima pembaruan yang jelas. Versi naskah, data analisis, skrip, dan alasan perubahan disimpan. Dalam penggunaan administratif, skor atau rekomendasi yang sudah dikeluarkan ditinjau ulang dan keputusan yang terdampak dikoreksi sesuai prosedur.

Jejak audit merekam siapa menjalankan tahap, kapan, versi alat, file input dan output, perubahan manual, persetujuan, dan alasan pengecualian. Catatan ini tidak harus mengarsipkan seluruh data sensitif atau setiap percakapan yang tidak relevan. Log harus cukup untuk menjawab pertanyaan audit tanpa memperbesar risiko privasi. Bila penggunaan model berlangsung pada layanan eksternal, tim mencatat kebijakan retensi serta menghindari menyimpan prompt berisi informasi personal pada repositori umum.

Sebuah prosedur eskalasi menjelaskan kapan pekerjaan dihentikan. Pemicu dapat berupa ketidakcocokan hasil yang tidak terjelaskan, kebocoran data, performa buruk pada kelompok tertentu, sumber referensi palsu, atau perubahan perilaku model setelah pembaruan vendor. Peneliti menghubungi pengelola data, statistikawan, komite etik, atau penanggung jawab institusi menurut jenis insidennya. Jangan meneruskan analisis hanya karena tenggat publikasi dekat. Ketidakpastian yang belum diselesaikan harus mengubah cakupan klaim atau menunda penggunaan sistem.

Setelah insiden, tim melakukan analisis akar penyebab. Apakah definisi kolom ambigu, data tidak seimbang, prompt mendorong klaim berlebihan, tidak ada pengujian baseline, atau satu orang mengendalikan pipeline tanpa peer review? Tindakan perbaikan dapat berupa kamus data lebih baik, validasi tambahan, batas akses, pelatihan, atau mengganti model dengan prosedur sederhana. Pelajaran dicatat dan dibagikan dalam bentuk yang melindungi subjek data. Budaya yang memungkinkan pelaporan kesalahan secara cepat memperkuat mutu ilmiah dan keselamatan penggunaan.

43. Analisis literatur dan ekstraksi bukti

Analisis data akademik dapat mencakup ekstraksi informasi dari artikel, laporan, registri, atau dokumen kebijakan. LLM mampu menemukan bagian yang tampak relevan, mengisi tabel karakteristik studi, dan membuat ringkasan awal. Namun ekstraksi bukan sekadar menemukan kalimat. Peneliti harus menentukan unit bukti, definisi variabel, waktu pengukuran, populasi, desain, outcome, serta hubungan antarhasil. Angka pada abstrak dapat berbeda dari tabel utama karena pembaruan, subkelompok, atau koreksi. Satu kalimat yang cocok dengan prompt belum tentu menjadi bukti yang tepat untuk pertanyaan sintesis.

Protokol ekstraksi disusun terlebih dahulu. Untuk setiap bidang, tulis definisi operasional, aturan jika tidak dilaporkan, cara menangani beberapa lengan studi, dan format satuan. Model dapat mengisi draf lalu menyertakan petunjuk halaman, tabel, atau kutipan yang mendukung nilai. Pemeriksa menelusuri kembali ke dokumen sumber dan menandai apakah nilai langsung dilaporkan, dihitung ulang, atau ditafsirkan. Jika model tidak dapat memberikan bukti lokasi, hasil dibiarkan belum terverifikasi. Informasi yang diekstrak tidak boleh ditransformasikan menjadi fakta yang lebih pasti daripada pelaporan aslinya.

Dalam sintesis bukti, risiko seleksi muncul bila sistem hanya menemukan dokumen yang mudah dibaca atau berbahasa dominan. OCR dapat gagal pada tabel, dokumen hasil scan, atau notasi ilmiah. Studi dengan hasil nol mungkin memiliki abstrak yang kurang menarik dan kurang terambil oleh pencarian semantik. Strategi pencarian tetap didokumentasikan dan dilakukan di sumber yang relevan, dengan proses screening yang ditetapkan. AI dapat membantu prioritisasi screening, namun tim memeriksa recall pada sampel dan menetapkan prosedur untuk dokumen yang terlewat.

Kinerja ekstraksi harus dinilai per variabel dan konsekuensi kesalahan. Salah satu angka ukuran sampel mungkin mudah dikoreksi; salah memilih outcome

primer dapat mengubah kesimpulan sintesis. Studi proof-of-concept pada sepuluh uji acak menemukan akurasi ekstraksi yang tinggi pada kumpulan elemen tertentu menggunakan model tertentu, tetapi desain dan ukuran tersebut tidak membuktikan kinerja universal di seluruh disiplin atau format artikel (Gartlehner et al., 2024). Karena itu, bukti awal mendukung percobaan terkontrol, bukan pelepasan pengawasan.

Sintesis naratif juga memerlukan keterlacakan. Setiap klaim yang menggabungkan studi harus terhubung ke referensi dan hasil spesifik; duplikasi artikel, publikasi turunan, atau laporan yang memakai sampel sama perlu dikenali. Model dapat merangkum setiap studi, tetapi sintesis membandingkan kualitas desain, populasi, ketidakpastian, dan heterogenitas. Peneliti menghindari menghitung jumlah studi yang "positif" seolah ukuran efek, serta tidak menafsirkan kesepakatan beberapa model sebagai validasi independen jika semuanya dilatih pada sumber dan pola serupa.

44. Merancang prompt dan batas interaksi yang dapat diaudit

Prompt bukan pengganti metode, tetapi merupakan bagian dari kondisi analisis ketika model generatif digunakan. Instruksi yang baik menyebut tugas, konteks yang boleh dipakai, bentuk output, kriteria kualitas, dan batas klaim. Untuk klasifikasi teks, misalnya, prompt dapat meminta label berdasarkan codebook, kutipan sumber, tingkat ketidakpastian, dan keputusan "tidak cukup informasi". Prompt tidak seharusnya meminta model mengarang penjelasan ketika data tidak mendukung. Versi prompt dicatat bersama model dan tanggal penggunaan agar tim mengetahui kondisi eksperimen.

Interaksi bertahap dapat meningkatkan kualitas jika setiap tahap mempunyai tujuan yang jelas: pertama memeriksa struktur data; kemudian menyarankan pemeriksaan; selanjutnya menghasilkan kode; lalu menjelaskan output. Namun percakapan panjang dapat mencampur perubahan asumsi dan membuat sulit mengenali dasar respons akhir. Tim menyimpan instruksi yang

menghasilkan artefak penting, serta mencatat revisi manual yang bermakna. Tidak perlu mengarsipkan percakapan yang tidak memengaruhi hasil, tetapi keputusan yang mengubah filter, kategori, atau interpretasi harus dapat ditelusuri.

Prompt dapat memengaruhi keluaran melalui cara bertanya. Instruksi "buktikan bahwa program berhasil" mendorong narasi satu arah, sedangkan instruksi yang netral meminta dukungan dan bukti yang bertentangan. Peneliti meninjau bias konfirmasi yang masuk lewat bahasa prompt. Untuk pekerjaan penting, tim dapat meminta analisis tandingan: temukan kasus yang tidak cocok, sebutkan asumsi, atau tunjukkan alasan kesimpulan mungkin salah. Pendekatan ini memperluas pemeriksaan, meskipun tetap tidak menjamin bahwa model menemukan semua kekeliruan.

Model perlu diberi data minimum yang dibutuhkan. Mengunggah seluruh dataset demi meminta bantuan pada satu error kode tidak proporsional. Contoh kecil yang telah disintesis atau disamarkan sering cukup untuk memperbaiki format dan logika. Jika kode memiliki variabel sensitif, nama variabel dapat diganti sebelum konsultasi, tetapi transformasi tersebut tidak menggantikan persetujuan layanan dan kebijakan institusi. Peneliti mengecek output karena model dapat mengulang kembali potongan data input dalam respons.

Kualitas interaksi juga dilihat dari kapan sistem menyatakan batas. Chatbot yang menyebut "saya tidak tahu" saat variabel tidak dijelaskan lebih aman daripada yang mengisi makna secara spekulatif. Peneliti dapat meminta sistem merinci asumsi yang digunakan, tetapi tetap menguji asumsi itu sendiri. Jika beberapa formulasi prompt menghasilkan kesimpulan berbeda, variasi tersebut menjadi temuan mengenai sensitivitas metode dan memerlukan prosedur yang stabil. Output terpilih tidak dipilih karena paling meyakinkan; kriteria pemilihan ditentukan sebelum melihat narasi final.

45. Batas tanggung jawab dan integritas akademik

Setiap hasil yang diterbitkan mempunyai pemilik tanggung jawab manusia. Sistem AI tidak dapat menjadi penulis yang memahami persetujuan, menanggapi koreksi, atau menanggung konsekuensi jika klaim salah. Peneliti yang tercantum sebagai penulis memastikan bahwa analisis dapat dijelaskan, sumber diverifikasi, angka dicocokkan dengan output, dan penggunaan AI mengikuti protokol serta aturan penerbit. Jika beberapa orang mengerjakan tahap berbeda, kontribusi mereka dicatat secara transparan. Tanggung jawab kolektif tidak boleh mengaburkan siapa yang memeriksa skrip, data, interpretasi, dan keputusan publikasi.

Integritas akademik tidak hanya tentang mengungkapkan nama alat. Pembaca perlu mengetahui bagian proses yang dibantu model, alasan penggunaannya, dampak terhadap data dan klaim, serta cara validasi. Pernyataan seperti "AI digunakan dalam penelitian" terlalu umum untuk memungkinkan penilaian. Uraian yang lebih berguna menyebut tugas spesifik: misalnya model membantu membuat draf kode untuk visualisasi distribusi, skrip direview dan dijalankan oleh penulis, kemudian angka pada naskah dicocokkan dengan tabel hasil. Pada pemrosesan teks, laporan menjelaskan ukuran data, prosedur anonimisasi, codebook, serta proporsi keluaran yang diperiksa manusia.

Ketika AI membantu menyusun kode atau kalimat hasil, pemeriksaan plagiarisme dan hak cipta masih diperlukan. Model dapat menghasilkan frasa mirip dengan sumber yang pernah dilihatnya, membuat sitasi palsu, atau mengutip artikel yang tidak menyatakan hal tersebut. Semua sumber dilacak ke naskah asli atau penerbit tepercaya; metadata DOI dan tahun diperiksa; kutipan diverifikasi pada konteks lengkap. Daftar pustaka yang disusun otomatis bukan bukti bahwa bacaan tersebut pernah diperiksa. Kesalahan bibliografis dapat merusak kredibilitas dan menyesatkan pembaca mengenai dasar argumen.

Peninjau dan editor juga memiliki kewajiban. Manuskrip yang tidak dipublikasikan dapat memuat gagasan baru, data sensitif, atau strategi yang belum diumumkan. Memasukkan naskah ke layanan generatif tanpa izin dapat melanggar kerahasiaan. Peninjau tidak mendelegasikan keputusan penerimaan kepada model; evaluasi tetap memakai kriteria ilmiah dan kebijakan penerbit. Penulis menghindari menempelkan komentar reviewer atau materi rahasia ke layanan yang tidak disetujui. Tata kelola integritas karena itu mencakup seluruh rantai produksi ilmu, dari data awal hingga telaah sejawat.

Transparansi yang efektif membantu pembaca menilai kekuatan bukti tanpa melebih-lebihkan kontribusi atau menyalahkan alat secara abstrak. Jika sistem membuat kesalahan, penulis bertanggung jawab memperbaiki analisis dan memberi koreksi. Jika model membantu secara rutin tetapi tidak memengaruhi hasil substantif, uraian tetap proporsional. Prinsip ini menjaga posisi AI sebagai sarana kerja dalam penyelidikan ilmiah, sementara otoritas metodologis dan akuntabilitas tetap melekat pada manusia yang merancang, memvalidasi, dan menggunakan hasilnya.

46. Kesimpulan

AI memperluas kemampuan analisis data akademik pada pemrograman, pengolahan teks, eksplorasi pola, prediksi, dan penyusunan ringkasan. Manfaat itu muncul ketika sistem ditempatkan dalam alur yang terdefinisi dan outputnya divalidasi. Risiko muncul ketika kecepatan disamakan dengan kebenaran, skor agregat dianggap mewakili semua kelompok, model prediktif ditafsirkan sebagai sebab, atau data sensitif dikirim tanpa pengamanan.

Analisis yang bertanggung jawab dimulai dengan tujuan ilmiah dan pertanyaan yang jelas. Peneliti memahami data, menetapkan metode sesuai inferensi, memeriksa kualitas, menguji kode, menilai ketidakpastian, membandingkan baseline, memeriksa fairness, dan menyimpan jejak

keputusan. AI boleh membantu menjelajah, tetapi peneliti harus mampu menjelaskan alasan hasil dipercaya. Untuk data akademik, perlindungan mahasiswa, karyawan, pasien, dan komunitas merupakan bagian dari mutu metodologis.

Artificial intelligence tidak menghapus keahlian manusia. Ia menggeser sebagian kerja dari penulisan sintaks dan pengelolaan rutin menuju perumusan pertanyaan, desain validasi, pemahaman konteks, dan pengawasan inferensi. Kemajuan dinilai bukan dari jumlah tugas yang diotomatisasi, tetapi dari apakah pengetahuan menjadi lebih akurat, berguna, terbuka, adil, dan dapat diperiksa.

Untuk mewujudkan tujuan itu, universitas perlu menjadikan kemampuan analisis sebagai praktik bersama. Mahasiswa dan peneliti belajar bukan hanya mengoperasikan alat, tetapi juga bertanya dari mana data berasal, asumsi apa yang dibuat, metode apa yang sesuai, siapa yang mungkin dirugikan, dan bukti apa yang cukup untuk menerima suatu klaim. Pelatihan dibangun melalui contoh nyata, peer review kode dan interpretasi, akses pada konsultasi statistik, serta kebijakan data yang mudah dipakai. Dosen memberi ruang untuk menjelaskan proses dan membela keputusan analitis, sehingga penggunaan AI tidak menggantikan pemahaman. Ketika pengetahuan metodologis tersebar luas, institusi dapat memperoleh manfaat dari otomasi tanpa memindahkan tanggung jawab kepada perangkat lunak. Dengan demikian, standar akademik berkembang seiring teknologi: keterbukaan tentang bantuan yang digunakan, ketelitian pemeriksaan, perlindungan data, dan keberanian memperbaiki kesimpulan ketika bukti berubah tetap menjadi dasar kepercayaan ilmiah.

Glossarium

Akurasi: Proporsi prediksi yang tepat menurut label acuan; maknanya bergantung pada distribusi kelas dan konsekuensi kesalahan.

Analitik pembelajaran: Pengumpulan, analisis, dan penggunaan data tentang pembelajar serta konteksnya untuk memahami dan memperbaiki pembelajaran.

Artificial intelligence (AI): Kumpulan sistem komputasi yang menjalankan tugas seperti prediksi, klasifikasi, optimasi, pemrosesan bahasa, atau generasi konten.

Bias data: Distorsi sistematis yang muncul dari cara data dikumpulkan, diukur, dipilih, dilabeli, atau digunakan.

Data leakage: Masuknya informasi dari data uji atau masa depan ke proses pelatihan sehingga kinerja terlihat lebih baik dari kondisi nyata.

Estimand: Besaran atau efek spesifik yang hendak diperkirakan dalam suatu penelitian.

Generative AI: AI yang menghasilkan konten baru seperti teks, kode, gambar, atau audio berdasarkan pola yang dipelajari.

Kalibrasi: Kesesuaian antara probabilitas prediksi model dan frekuensi hasil aktual.

Machine learning: Metode yang mempelajari pola dari data untuk prediksi, klasifikasi, estimasi, atau representasi.

Model drift: Perubahan kinerja ketika hubungan atau distribusi data berubah dari kondisi saat model dikembangkan.

Nilai hilang: Nilai variabel yang tidak tercatat atau tidak tersedia untuk sebagian observasi.

Overfitting: Keadaan ketika model terlalu menyesuaikan diri dengan data latih dan gagal menggeneralisasi.

Pipeline analisis: Rangkaian proses dari input data sampai hasil dan laporan, termasuk transformasi serta kontrol kualitas.

Reproduksibilitas: Kemampuan menghasilkan kembali hasil dengan data dan prosedur yang sama atau terdokumentasi.

Synthetic data: Data yang dihasilkan model atau simulasi untuk meniru karakteristik tertentu dari data empiris.

Daftar Pustaka

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. Computers and Education: Artificial Intelligence, 6, Article 100215. https://doi.org/10.1016/j.caeai.2024.100215

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, *6*, Article 100215. https://doi.org/10.1016/j.caeai.2024.100215

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Combrinck, C. (2024). A tutorial for integrating generative AI in mixed methods data analysis. Discover Education, 3, Article 116. https://doi.org/10.1007/s44217-024-00214-7

Combrinck, C. (2024). A tutorial for integrating generative AI in mixed methods data analysis. *Discover Education*, *3*, Article 116. https://doi.org/10.1007/s44217-024-00214-7

Gartlehner, G., Kahwati, L., Hilscher, R., Thomas, I., Kugley, S., & Crotty, K. (2024). Data extraction for evidence synthesis using a large language model: A proof-of-concept study. Research Synthesis Methods, 15(4), 576–589. https://doi.org/10.1002/jrsm.1710

Gartlehner, G., Kahwati, L., Hilscher, R., Thomas, I., Kugley, S., & Crotty, K. (2024). Data extraction for evidence synthesis using a large language model: A proof-of-concept study. *Research Synthesis Methods*, *15(4)*, 576–589. https://doi.org/10.1002/jrsm.1710

Ioannidis, J. P. A. (2005). Why most published research findings are false. PLOS Medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2(8)*, e124. https://doi.org/10.1371/journal.pmed.0020124

Jalali, M. S., & Akhavan, A. (2024). Integrating AI language models in qualitative research: Replicating interview data analysis with ChatGPT. System Dynamics Review, 40(3), e1772. https://doi.org/10.1002/sdr.1772

Jalali, M. S., & Akhavan, A. (2024). Integrating AI language models in qualitative research: Replicating interview data analysis with ChatGPT. *System Dynamics Review*, *40(3)*, e1772. https://doi.org/10.1002/sdr.1772

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns, 4(9), 100804. https://doi.org/10.1016/j.patter.2023.100804Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based

science. *Patterns*, *4(9)*, 100804. https://doi.org/10.1016/j.patter.2023.100804

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Nguyen-Trung, K. (2025). ChatGPT in thematic analysis: Can AI become a research assistant in qualitative research? Quality & Quantity, 59, 4945–4978. https://doi.org/10.1007/s11135-025-02165-z

Nguyen-Trung, K. (2025). ChatGPT in thematic analysis: Can AI become a research assistant in qualitative research? *Quality & Quantity*, *59*, 4945–4978. https://doi.org/10.1007/s11135-025-02165-z

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366(6464)*, 447–453. https://doi.org/10.1126/science.aax2342

Prandner, D., Wetzelhütter, D., & Hese, S. (2025). ChatGPT as a data analyst: An exploratory study on AI-supported quantitative data analysis in empirical research. Frontiers in Education, 9, Article 1417900. https://doi.org/10.3389/feduc.2024.1417900

Prandner, D., Wetzelhütter, D., & Hese, S. (2025). ChatGPT as a data analyst: An exploratory study on AI-supported quantitative data analysis in empirical research. *Frontiers in Education*, *9*, Article 1417900. https://doi.org/10.3389/feduc.2024.1417900

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding

Principles for scientific data management and stewardship. Scientific Data, 3, Article 160018. https://doi.org/10.1038/sdata.2016 .8

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, Article 160018. https://doi.org/10.1038/sdata.2016.18