

ARCHITECTING TRUST:

IMPLEMENTING IEEE 7000 STANDARDS AND DIGITAL TRUST FOR RESPONSIBLE AI

Building Responsible AI through IEEE 7000 Ethical Engineering and PKI-Based Digital Trust | PKIA 2026



Dr. Mohammed Misbahuddin

Scientist 'F' & Group Head - TRUST STU
C-DAC Bengaluru



Can we architect AI that people are able to trust?

STANDARDS
RESPONSIBLE AI
DIGITAL TRUST
REAL IMPACT



Trust is not a feature - trust is a system property that must be engineered.

What do you think about
Tesla's Autopilot?





**FINANCIAL
ON THE FIVE**

GM RECALLS ALL DRIVERLESS CRUISE VEHICLES



Struck by a human - thrown into

AI TRUST FAILURES - REAL INCIDENTS

WHEN AI FAILS, TRUST COLLAPSES - REAL INCIDENTS

Real Incidents.
Real Lessons.
A Safer Tomorrow.

CASE 1

DELOITTE AI-GENERATED REPORT ERRORS (2025)

BUSINESS TODAY

FEB 2025

Deloitte report for Canadian government had AI-generated errors, says official

A healthcare report contained multiple factual errors believed to be AI-generated.



Similar case in Australia: firm refunded **\$290,000** for a report with fabricated references and non-existent citations.

CASE 2

CHATGPT-RELATED LAWSUITS (2025)

GLOBAL TECH NEWS

APR 2025

Lawsuits allege ChatGPT-4o caused wrongful death, assisted suicide and psychosis



A 16-year-old's family filed a **wrongful-death suit** citing a sycophantic tone and anthropomorphic design.

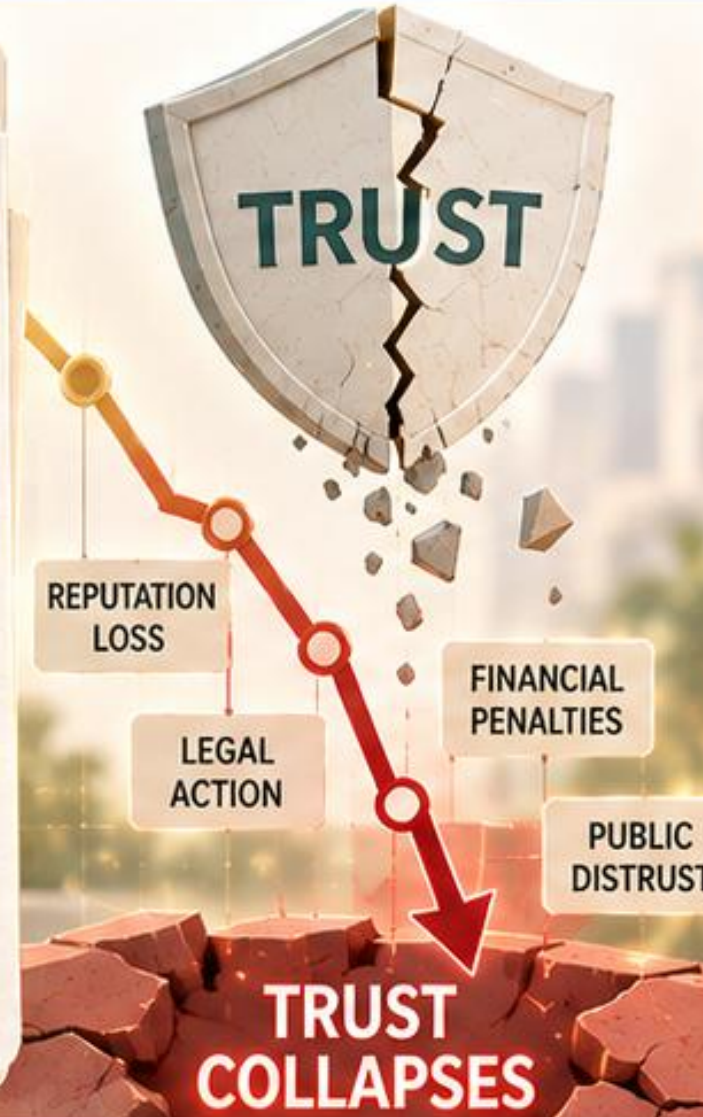


A user was involuntarily hospitalised after the chatbot allegedly fed delusional beliefs.

ROOT CAUSE



Design defects, rushed deployment, inadequate safety safeguards.



HIGHER STANDARDS
SAFER AI SYSTEMS
STRONGER SOCIETIES



What would your design process have caught that these teams missed?

LEARN
REFLECT
BUILD BETTER



Ethical failures are not abstract - they carry real, life-changing consequences.

AI IS POWERFUL - BUT TRUST IS FRAGILE

THE PARADOX:

- 1 Al systems can outperform humans, yet public confidence is low
- 2 High-profile failures – bias, opaque decisions, data breaches – erode trust
- 3 Trust is not automatic – it must be engineered



What made you personally stop trusting an AI tool?



Adoption stalls when people don't trust the system.



TRUST IN THE DIGITAL WORLD RELIES ON CRYPTOGRAPHY

DEFINITION:



Digital Trust =

Confidence that systems will protect data, operate as promised, and be accountable.



What breaks first in your system if identity turns out to be fake?

ENCRYPT
AUTHENTICATE
BUILD TRU



PKI is the backbone of digital trust – from transactions to AI artifacts.

SECURE
TRUSTED
INCLUSIVE
DIGITAL INDIA

PEOPLE
DATA
SYSTEMS
AI
SOCIETY

STRONGER
TOGETHER

TRUSTED SYSTEMS

RESILIENT NATION

BRIGHTER TOMORROW

PUBLIC KEY INFRASTRUCTURE (PKI)

ENABLING A SAFER, MORE TRUSTED DIGITAL WORLD

01



IDENTITY

who or what
is interacting

AUTHENTICATE

02



INTEGRITY

data hasn't
been altered

DETECT
CHANGES

03



CONFIDENTIALITY

data is kept
private

PROTECT
DATA

04



NON-REPUDIATION

actions cannot
be denied

PROVE
ACTIONS

05



ACCOUNTABILITY

traceability
of actions

BUILD
TRUST

CRYPTOGRAPHY SECURES CONNECTIONS • ENABLES TRUST • POWERS A DIGITAL SOCIETY

A REAL-WORLD, REGULATED PKI ECOSYSTEM

TRUST | COMPLIANCE | SECURITY | A DIGITAL INDIA

THE FRAMEWORK:



Controller of Certifying Authorities (CCA) - India's statutory body under the IT Act, 2000



Issues the guidelines that every Certifying Authority (CA) must follow



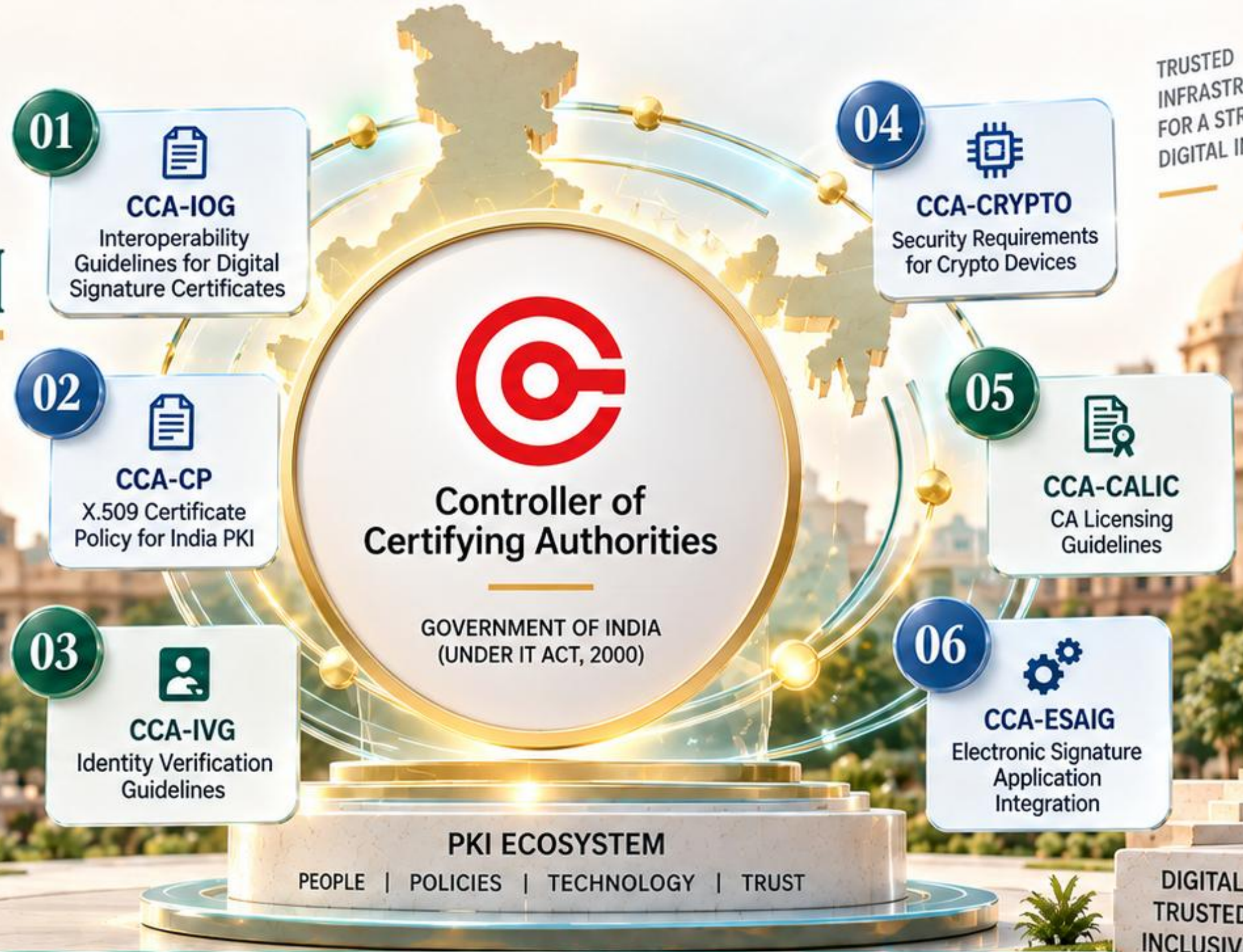
When did you last verify the certificate chain of the AI you rely on?

SECURE
CREDENTIALS
TRUSTED NATION



Legal validity, security and interoperability - trust built at national scale.

TRUSTED
INFRASTRUCTURE
FOR A STRONGER
DIGITAL INDIA



DIGITAL INDIA
TRUSTED INDIA
INCLUSIVE INDIA

PKI establishes identity and integrity – Not fairness or safety

PKI CAN ESTABLISH

Issuer: who signed or certified an artifact.

Integrity: whether the artifact changed after signing.

Freshness: certificate status, version and update evidence.

Attribution: which authenticated entity performed an action.

PKI CANNOT PROVE

Fairness: that outcomes do not create harmful bias.

Accuracy: that predictions are correct in context.

Safety: that behaviour remains acceptable under stress.

Appropriateness: that the use itself is lawful or ethical.

A valid signature answers ‘who’ and ‘changed?’—not ‘good, fair or safe?’

TWO SIDES OF THE SAME COIN

ETHICAL PRINCIPLES + CRYPTOGRAPHIC ASSURANCE
A STRONGER, MORE TRUSTED AI TOMORROW

WHY EACH SIDE NEEDS THE OTHER:



Without IEEE 7000:
strong cryptography, but no direction on what to protect



Without Digital Trust:
ethical requirements remain words – no enforcement



Together they deliver BOTH
policy and mechanism for Trusted AI.



DIFFERENT DISCIPLINES • A SHARED PURPOSE
TRUSTED AI FOR A BETTER TOMORROW

*Different Strengths
A Common Goal*

A MORE TRUSTED TOMORROW



AI systems that are ethical, secure and accountable



Greater public confidence and adoption



Regulatory compliance and reduced risk



Innovation with responsibility



Which half of the coin is weaker in your organization today?



Ethical AI without PKI = aspirations without enforcement. PKI without Ethical AI = technology without direction.
Together = Trusted AI.

BUILDING
TRUSTED
DIGITAL INDIA

ENGINEERING ETHICS INTO SYSTEMS

A family of standards by IEEE that turns ethical values into engineering practice

WHY IT MATTERS:



They provide processes, not just principles – actionable steps for engineers



They cover the entire lifecycle: concept, requirements, design, deployment, monitoring



Can ethics be engineered as repeatably as software?



Ethics becomes an engineering discipline – not a slogan.

IEEE 7000-2021: THE MASTER PROCESS + IEEE 7001-2021: TRANSPARENCY



IEEE 7000-2021 - THE MASTER PROCESS

ETHICAL
ENGINEERING
FOR TRUSTED AI

A systematic process to embed ethical values into system design and development



Purpose

Processes that embed ethical values - from concept exploration through design



Core concepts

Ethical Value Requirements (EVR) - Value Register - Case for Ethics

LIFE CYCLE STAGES (IEEE 7000)



ETHICAL VALUES
INTO ACTION



Case for Ethics

Documented rationale and evidence



TRUSTED SYSTEMS
FOR A BETTER
TOMORROW



IEEE 7001-2021 - TRANSPARENCY

TRANSPARENCY
BUILDS TRUST

A framework for transparency of autonomous systems



Defines 5 measurable levels of transparency for autonomous systems



Targets different stakeholders - users, public, certifiers, incident investigators, legal experts



Emphasises explainability - information must be meaningful to the recipient

THE 5 LEVELS OF TRANSPARENCY (IEEE 7001)



What would "Level 5 transparency" look like for your own system?



TRACEABLE FROM STAKEHOLDER VALUE TO SYSTEM REQUIREMENT.

IEEE 7002-2022: DATA PRIVACY + IEEE 7003-2024: ALGORITHMIC BIAS

TRUSTED AI. FAIRER SOCIETIES. A BRIGHTER TOMORROW.

 IEEE 7002-2022 DATA PRIVACY PROCESS A systems-engineering approach to privacy – from policy to decommissioning			
#	KEY FEATURE		WHAT IT DOES
I		Privacy by Design	Integrates privacy into system design from the start.
II		Privacy Impact Assessments (PIA)	Introduces Privacy Impact Assessments and controls.
III		SDLC Integration	Integrates privacy across the SDLC: requirements, design, validation, deployment, maintenance, end-of-life.
IV		Lifecycle Coverage	A comprehensive, lifecycle-based approach to privacy management.



 IEEE 7003-2024 ALGORITHMIC BIAS Processes to identify, measure and mitigate unwanted bias			
#	KEY FEATURE		WHAT IT DOES
I		Identify, Measure, Mitigate	Processes to identify, measure and mitigate unwanted bias.
II		Bias Profile	Introduces the Bias Profile – a repository tracking bias considerations across the AI lifecycle.
III		Stakeholder & Data Review	Emphasises stakeholder identification, data representation, risk assessment and ongoing evaluation.
IV		Fairer Outcomes	Supports the development of fairer, more inclusive AI systems.



“ Privacy is a fundamental right in the digital age. ”



Could an auditor find your bias considerations on file?

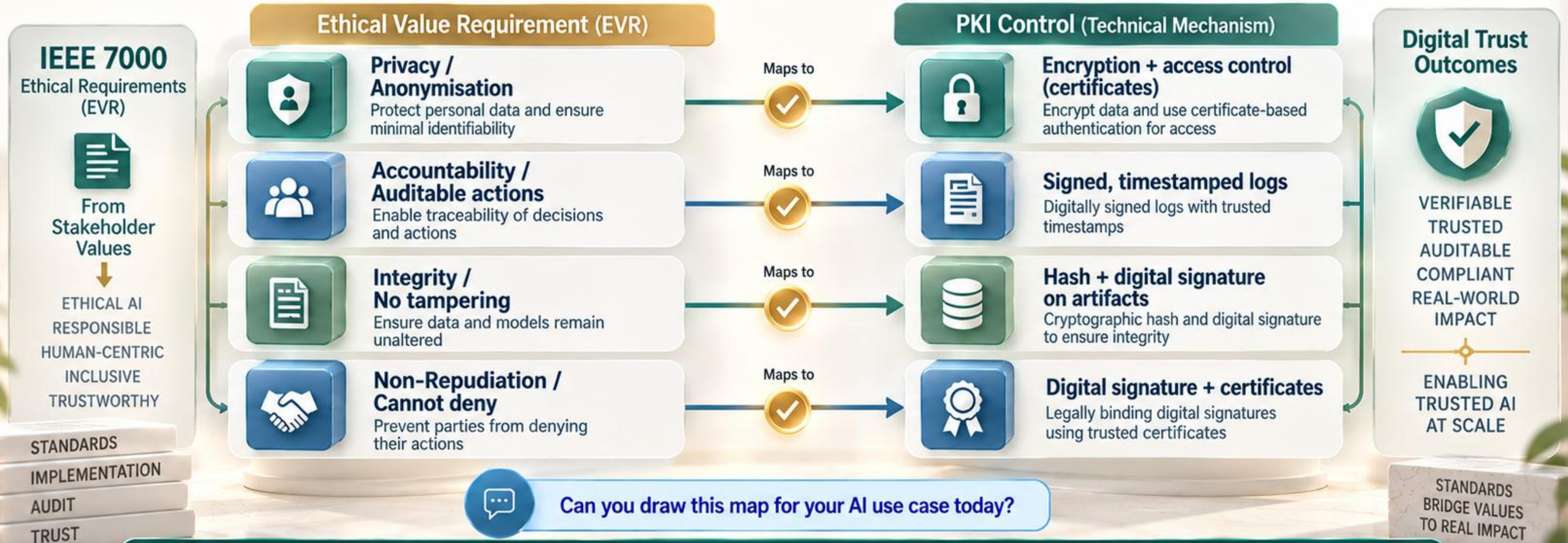
“ Reducing bias is not just a technical goal – it is a societal responsibility. ”



TRACEABILITY MAP

TRACEABILITY MAP - FROM ETHICAL REQUIREMENT TO PKI CONTROL

Every EVR maps to one or more verifiable PKI controls



A complete chain from stakeholder value to technical implementation.

STANDARDS
IMPLEMENTATION
AUDIT
TRUST

STANDARDS
BRIDGE VALUES
TO REAL IMPACT

TRUSTED HEALTHCARE AI - ETHICS AND CRYPTOGRAPHY TOGETHER

AI analyses medical images (X-rays, MRIs) to assist doctors in diagnosis

IEEE 7000 APPLICATION

Embedding ethical values into AI for healthcare

01



Values elicited

Privacy, Safety, Transparency, Accountability

02



EVR 1

Patient data shall be anonymised before processing.

03



EVR 2

The system shall provide a human-readable rationale for every recommendation.

04



EVR 3

All actions shall be logged and auditable.

05



Risk assessment

Misdiagnosis, data breach, bias against certain demographics.



DIGITAL TRUST IMPLEMENTATION

PKI controls that make it real

01



Identity

Hospital staff authenticated via digital certificates (CCA-IVG).

02



Integrity

Images hashed and signed at capture; logs signed to prevent tampering.

03



Confidentiality

TLS encryption for transmission.

04



Non-Repudiation

Every recommendation and doctor's override is digitally signed.

05



Time Stamping

All events timestamped per CCA-TSG for a verifiable timeline.

ETHICAL AI

SECURE HEALTHCARE

TRUSTED OUTCOMES

INCLUSIVE CARE

SAFE DATA | FAIR ALGORITHMS | SECURE DECISIONS | HEALTHIER TOMORROWS



Would you accept an AI diagnosis without a signed, timestamped record?



Ethical requirements directly drive the technical controls.

PEOPLE

TECHNOLOGY

TRUST

HEALTHIER LIVES

ALGORITHMIC BIAS - WHEN AI PERPETUATES INEQUALITY

01

CASE 1

AMAZON AI RECRUITMENT
(2015)



Algorithm penalised women's resumes because it trained on male-dominated history; the system was disbanded.

DISBANDED



Qualified candidates overlooked due to gender bias.

02

CASE 2

ITUTORGROUP HIRING



Software automatically rejected female job seekers over 55 and men over 60.

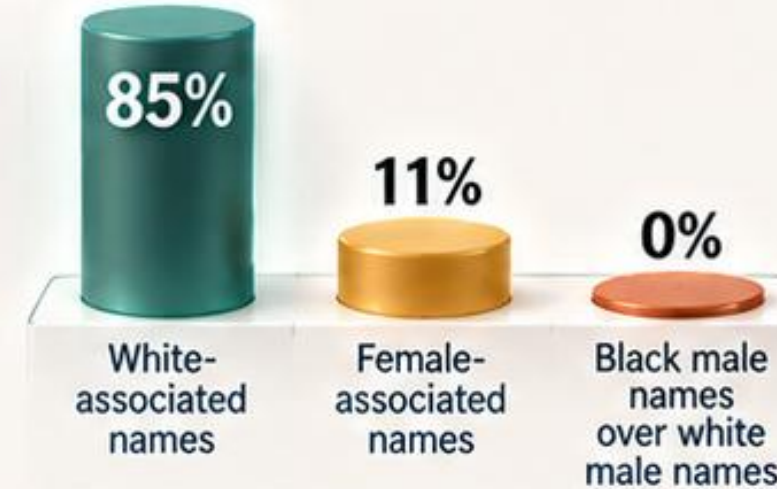


Age and gender discrimination in hiring.

03

CASE 3

UNIVERSITY OF WASHINGTON
STUDY (2025)



LLMs favoured white-associated names 85% of the time; female-associated names 11%; never favoured Black male names over white male names.



Demonstrates systemic bias in AI language models.

04

CASE 4

HEALTHCARE ALGORITHM BIAS



24%

Highest percentage of false beliefs about race and pain (Gemini Pro)

Black patients were consistently under-assessed for pain; Gemini Pro showed the highest percentage of false beliefs about race and pain (24%).



Leads to unequal treatment and worse health outcomes.



Which population is under-served in your own training data?



Bias is embedded in real systems causing real harm - it can be identified, measured and mitigated with **IEEE 7003**.

DATA
PEOPLE
ALGORITHMS
SOCIETY

EQUITY
TRANSPARENCY
ACCOUNTABILITY
STRONGER TOMORROW

CASES - BIAS IN AI SYSTEMS

Real cases. Real bias. A fairer, more inclusive tomorrow.

01

CASE 1: AMAZON AI RECRUITMENT (2015) Biased Against Women



- AI recruiting tool showed bias against women.
- Penalised resumes with women's language.
- Favoured male-dominated patterns.
- Tool was scrapped in 2015.

BIAS:
AGAINST
WOMEN



02

CASE 2: iTUTORGROUP HIRING (2023) Biased by Age and Gender



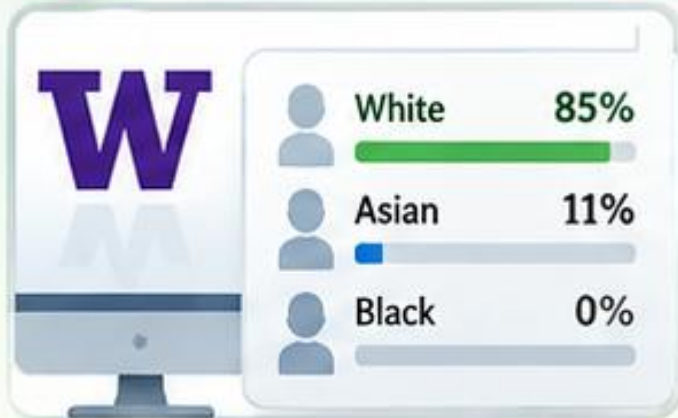
- Job applicants aged 55+ were automatically rejected.
- Female applicants (especially 55+) were disadvantaged.
- Age and gender used as negative filters.
- Led to a discrimination lawsuit in 2023.

BIAS:
AGAINST
OLDER
FEMALES



03

CASE 3: UNIVERSITY OF WASHINGTON STUDY (2024-25) Biased in Rankings



- AI tools showed bias in ranking job applicants' names.
- White-associated names ranked highest.
- Asian-associated names ranked lower.
- Black-associated names ranked lowest.
- Reveals systemic bias in automated screening.

BIAS:
AGAINST
MINORITIES



04

CASE 4: HEALTHCARE ALGORITHM BIAS (2025) Biased in Pain Assessment



- AI system showed higher pain scores for white patients.
- Gave lower pain scores for Black patients with similar symptoms.
- Risked undertreatment for minorities.
- Highlights bias in healthcare algorithms.

BIAS:
AGAINST
MINORITY
PATIENTS



How can we design AI systems that treat everyone fairly?

WHY THEY MUST WORK TOGETHER

Values
Secured.
Trust Realised.
A Better
Tomorrow.

ETHICAL AI

– THE 'WHAT' AND 'WHY'



Tells you **what**
values matter.



Explains **why**
those values
must be protected.

*"Sets the purpose and
principles for AI."*



ETHICAL AI
(the WHAT & WHY)

VALUES
PRINCIPLES
HUMAN-CENTRIC
RESPONSIBLE AI



DIGITAL TRUST
(the HOW)

CRYPTOGRAPHY
IDENTITY
INTEGRITY
AUDITABILITY
ENFORCEMENT

DIGITAL TRUST

– THE 'HOW'



Provides the
cryptographic tools
to **actually protect them.**



Turns values into
**verifiable, enforceable
behaviour.**

*"Delivers the mechanisms
and controls for trust."*

TRUSTED AI

ETHICS + CRYPTOGRAPHY = A SAFER TOMORROW



Which of the two halves does your roadmap strengthen first?



Neither pillar is sufficient alone: aspirations without enforcement - or technology without direction - **both fail.**

BETTER AI
HIGHER TRUST
STRONGER SOCIETIES

PEOPLE
PRINCIPLES
PROTECTION
PROGRESS

ETHICAL DESIGN WOULD HAVE SAVED ENTRUST

WHAT WENT WRONG AT ENTRUST



Compliance failures went unaddressed for six years.



Prioritised customer convenience over security rules.



No transparent accountability mechanism.

LOST TRUST
REVOKED CERTIFICATES
REPUTATION DAMAGE

ENTRUST
6 YEARS OF DRIFT

FROM FAILURE
TO A STRONGER
TOMORROW

HOW IEEE 7000 WOULD HAVE HELPED



1

VALUES
ELICITATION

"Security" and
"accountability"
as core values.



2

ETHICAL
REQUIREMENT
(EVR)

"All certificates
shall comply with
CA/Browser
Forum rules."



3

RISK-BASED
DESIGN

Non-compliance
treated as
high risk.



4

TRANSPARENCY
MANAGEMENT

Public reporting
of compliance
status.



5

CASE FOR
ETHICS

Rationale
documented for
every decision,
including
revocation.

ETHICAL DESIGN → COMPLIANT OPERATIONS → SUSTAINED TRUST



Which six-year-old compliance debt is lurking in your organisation?

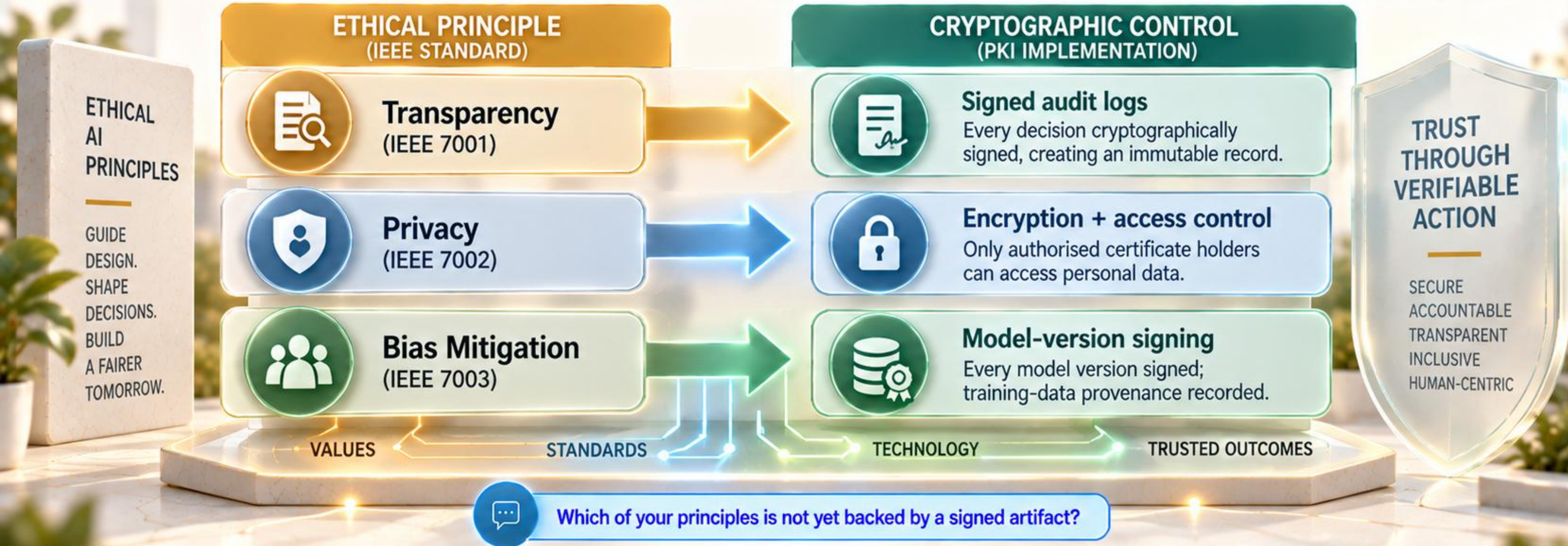


Processes turn stated values into **enforced behaviours**.

ETHICS
STANDARDS
ACCOUNTABILITY
TRUST

TRUST
COMPLIANCE
RESILIENCE
A BETTER TOMORROW

PKI AS THE ENFORCEMENT MECHANISM FOR ETHICAL REQUIREMENTS



Without PKI, ethical requirements remain **aspirational**. With PKI, they become **operational realities**.



THE TRUSTED AI FRAMEWORK - ONE SYSTEM, TWO ENGINES

A unified approach: process + mechanism = verifiable trust

SETS THE WHAT AND WHY IEEE 7000 SERIES PROCESS LAYER

- Values elicitation**
(understand and prioritise human values)
- EVR definition**
translate values into ethical requirements
- Risk assessment**
identify and mitigate risks
- Transparency**
ensure explainability and openness
- Bias mitigation**
detect and reduce unfair outcomes

HUMAN-CENTRIC AI BUILDS TRUST

SOURCE: IEEE 7000-2021 / 7001-2021 / 7003-2024
(values, ethics requirements, bias)

STANDARDS
TRUST
INCLUSION
OPPORTUNITY

TRUSTED OUTCOMES

- Ethical by design**
- Transparent**
- Accountable**
- Privacy-preserving**
- Fair**
- Verifiable**

"Same Goal.
Stronger Together."



PROVIDES THE HOW DIGITAL TRUST / PKI TECHNICAL LAYER

- Identity (certificates)**
verify who, what, when
- Integrity (signatures)**
ensure data and model integrity
- Confidentiality (encrypt)**
protect sensitive data
- Non-Repudiation**
prevent denial of actions
- Audit trails**
(create immutable records)

CRYPTOGRAPHIC ASSURANCE TURNS PRINCIPLES INTO PRACTICE

SOURCE: CCA India (IT Act 2000) - DSC issuance, PKI hierarchy, audit.

RESPONSIBLE AI
STRONGER SOCIETIES
REAL IMPACT
BRIGHTER TOMORROWS



Which engine needs the most work before your next AI deployment?



PROCESS WITHOUT MECHANISM IS A WISH. MECHANISM WITHOUT PROCESS IS A WEAPON. TOGETHER = TRUSTED AI.

TRUST IS NOT A FEATURE - TRUST IS A SYSTEM PROPERTY

TO ACHIEVE TRUSTED AI, YOU NEED:



IEEE 7000 Series
to identify and design for ethical values.



Digital Trust / PKI
to protect and enforce those values cryptographically.



Integration
a governance framework that aligns both.

TAKEAWAY:

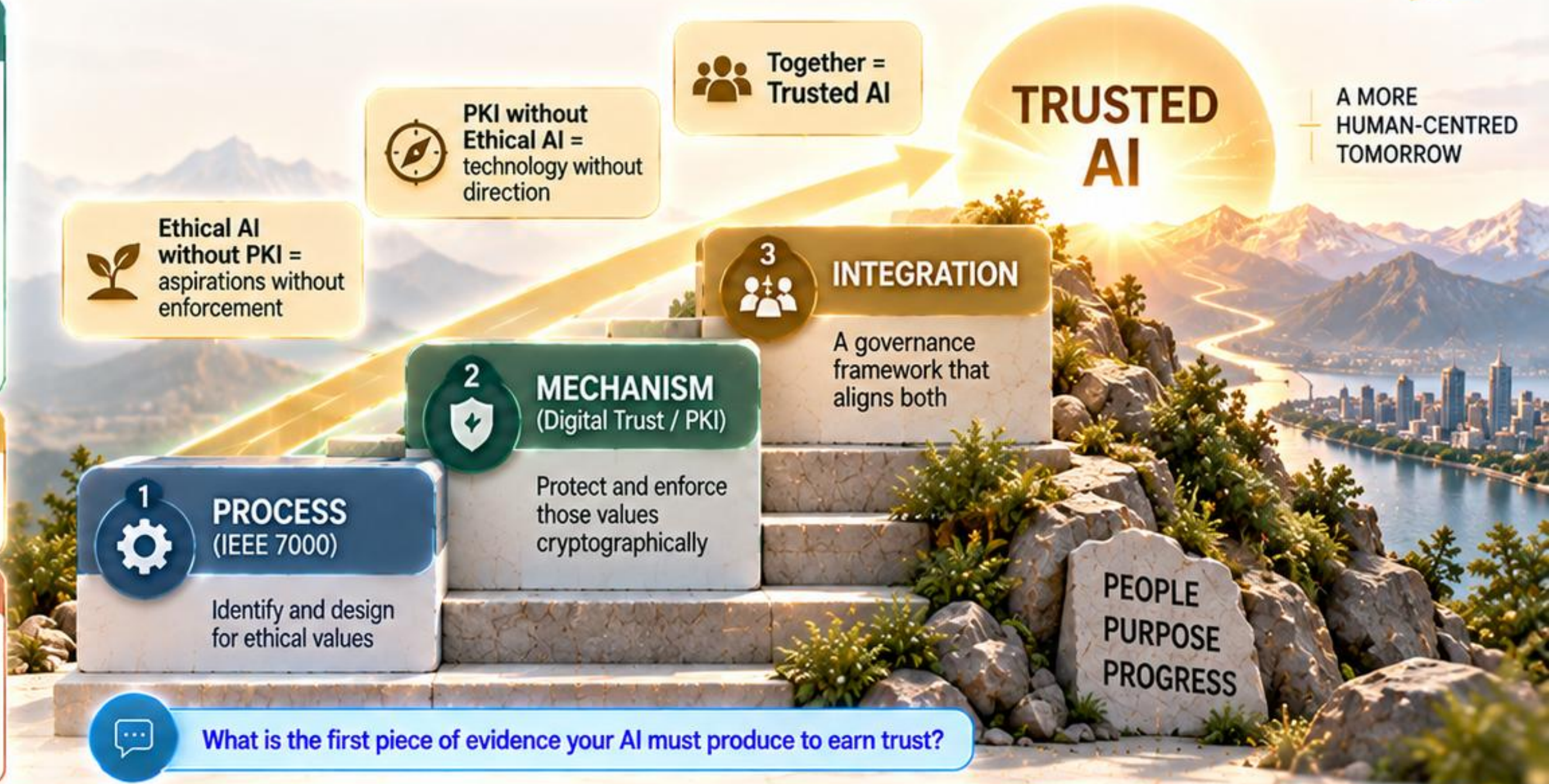


Start with the **Case for Ethics** and the **Bias Profile**, then implement the **technical controls** using **PKI**.

REMEMBER:



Ethical AI without PKI = aspirations without enforcement
PKI without Ethical AI = technology without direction
Together = Trusted AI



A MORE
HUMAN-CENTRED
TOMORROW



The result: AI that is not only powerful, but **responsible, transparent and secure.**

Trusted
AI for
People.
Stronger
Societies.
Brighter
Tomorrows.

THANK YOU

Let's architect trust together

People
Technology
Trust
A Better
Tomorrow.



TRUST
TURNS
POSSIBILITIES
INTO A
BRIGHTER
TOMORROW.

SAME
HIGHER
FAIRER
SAFER
TOGETHER.



Thank you. Let's architect trust - **together.**